



SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

COLLATED WRITTEN EVIDENCE VOLUME

Contents

10x Future Technology – Written evidence (AIC0024).....	1
The Academy of Medical Sciences – Written evidence (AIC0210).....	7
Accenture UK Limited – Written evidence (AIC0191).....	14
Advanced Marine Innovation Technology Subsea Ltd – Written evidence (AIC0038).....	25
Agents, Interaction and Complexity (AIC) group, University of Southampton – Written evidence (AIC0115)	28
AGISI.org – Written evidence (AIC0184).....	30
The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB) – Written evidence (AIC0086)	39
The AI Initiative, The Future Society at Harvard Kennedy School – Written evidence (AIC0209).....	45
The Alan Turing Institute – Written evidence (AIC0139).....	50
Mr Jaafar Almusaad and Mr Philip Bree – Written evidence (AIC0039).....	61
Amnesty International – Written evidence (AIC0180).....	64
Dr Sally Applin – Written evidence (AIC0172)	76
Arm – Written evidence (AIC0083)	79
Article 19 – Written evidence (AIC0129)	84
The Association of Medical Research Charities (AMRC) and the Wellcome Trust – Written evidence (AIC0202)	92
Dr Shahar Avin, Martina Kunz, Andrew Ware, Dr Simon Beard and Dr Seán Ó hÉigeartaigh – Written evidence (AIC0150)	93
Baker McKenzie – Written evidence (AIC0111).....	94
Balderton Capital (UK) LLP – Written evidence (AIC0232)	98
Professor Andrew Basden – Written evidence (AIC0195)	102
BBC – Written evidence (AIC0204)	115
BCS, The Chartered Institute for IT – Written evidence (AIC0049).....	123
Dr Simon Beard, Dr Seán Ó hÉigeartaigh, Dr Shahar Avin, Martina Kunz and Andrew Ware – Written evidence (AIC0150)	130
Miles Berry – Written evidence (AIC0247)	137
Big Brother Watch – Written evidence (AIC0154)	140

Big Innovation Centre – Written evidence (AIC0119)	149
Bikal – Written evidence (AIC0052)	161
Dr Richard Billingsley – Written evidence (AIC0201)	164
BioCentre – Written evidence (AIC0169)	175
Bioss International Ltd – Written evidence (AIC0033)	182
Dr Andrew Blick – Written evidence (AIC0064)	189
Dr Paula Boddington – Written evidence (AIC0067)	195
Michael Borgeaud – Written evidence (AIC0233)	199
Braintree – Written evidence (AIC0074)	205
Mr Philip Bree and Mr Jaafar Almusaad – Written evidence (AIC0039)	214
Bristows LLP – Written evidence (AIC0097)	215
The British Academy – Written evidence (AIC0213)	224
The British Institute of Facilities Management (BIFM) – Written evidence (AIC0205)	230
British Standards Institution – Written evidence (AIC0165)	241
British Standards Institution – Supplementary written evidence (AIC0231)	246
BSA The Software Alliance – Written evidence (AIC0153)	248
Dr Aysegul Bugra, Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis and Dr Valentina Rita Scotti – Written evidence (AIC0051)	257
Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams and Professor Robert Fisher – Written evidence (AIC0029)	265
Dr Mercedes Bunz – Written evidence (AIC0048)	266
Eur. Ing. David Burden and Professor Maggi Savin-Baden – Written evidence (AIC0061)	270
Michael Butterworth, Ms Joanna Goodman, Dr Paresh Kathrani, Dr Steven Cranfield and Chrissie Lightfoot – Written evidence (AIC0104)	279
Cancer Research UK – Written evidence (AIC0219)	280
Capco – Written evidence (AIC0071)	287
CBI – Written evidence (AIC0114)	298
Center for Data Innovation – Written evidence (AIC0043)	309
Centre for Health Economics University of York – Written evidence (AIC0242)	318
Centre for Public Impact – Written evidence (AIC0173)	321
Centre for the Study of Existential Risk and the Leverhulme Centre for the Future of Intelligence – Written evidence (AIC0237)	323

Centre for the Study of Existential Risk and the Leverhulme Centre for the Future of Intelligence – Supplementary written evidence (AIC0239)	324
CENTURY Tech – Written evidence (AIC0084)	325
Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis, Dr Valentina Rita Scotti and Dr Aysegul Bugra – Written evidence (AIC0051)	328
Charities Aid Foundation – Written evidence (AIC0042).....	329
Mr Thomas Cheney – Written evidence (AIC0098).....	336
Dr Esyin Chew – Written evidence (AIC0166)	341
Children’s Commissioner for England – Written evidence (AIC0123)	347
CIFAR – Written evidence (AIC0136)	351
Donald Clerk – Written evidence (AIC0022).....	357
CognitionX – Written evidence (AIC0170).....	360
Cognitive Finance Group – Written evidence (AIC0010).....	367
Competition and Markets Authority – Written evidence (AIC0245).....	371
Contact Centre Systems Ltd. – Written evidence (AIC0032).....	374
Cooley (UK) LLP – Written evidence (AIC0217)	379
Dr Steven Cranfield, Chrissie Lightfoot, Michael Butterworth, Ms Joanna Goodman and Dr Paresh Kathrani – Written evidence (AIC0104).....	385
Will Crosthwait – Written evidence (AIC0094).....	386
Darktrace – Written evidence (AIC0243)	392
Data & Society Research Institute – Written evidence (AIC0221)	400
Mr Graeme Davis – Written evidence (AIC0054).....	408
Deep Learning Partnership – Written evidence (AIC0027).....	412
Deep Science Ventures – Written evidence (AIC0167).....	421
DeepMind – Written evidence (AIC0234).....	427
Deloitte – Written evidence (AIC0075).....	444
Department of Computer Science University of Bath – Written evidence (AIC0099)	450
Department of Computer Science, University of Liverpool – Written evidence (AIC0192)	456
Digital Catapult – Written evidence (AIC0175)	464
Doteveryone – Written evidence (AIC0148)	473
Reverend Dr Lyndon Drake – Written evidence (AIC0108).....	478
Richard Ebley – Written evidence (AIC0026).....	481
The Economic Singularity Supper Club – Written evidence (AIC0058)	482
Professor Lilian Edwards – Written evidence (AIC0161).....	486

Electronic Frontier Foundation – Written evidence (AIC0199).....	491
Dr Julian Estevez – Written evidence (AIC0021).....	499
euRobotics Topics Group on ‘Ethical, Legal and Socio-economic issues’ – Written evidence (AIC0189).....	503
Faethm Pty Ltd – Written evidence (AIC0141).....	507
Family Law Partners – Written evidence (AIC0089)	512
Dr Jerry Fishenden – Written evidence (AIC0028)	518
Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams – Written evidence (AIC0029)	527
Dr Malcolm Fisk – Written evidence (AIC0012).....	535
Five AI Ltd – Written evidence (AIC0128).....	541
Foundation for Responsible Robotics – Written evidence (AIC0188)	547
Professor John Fox – Written evidence (AIC0076)	551
Laurence Freeman and Fabia Howard-Smith – Written evidence (AIC0147)	558
Fujitsu – Written evidence (AIC0120)	559
Future Advocacy – Written evidence (AIC0121).....	569
Future Intelligence – Written evidence (AIC0216).....	579
Future of Humanity Institute – Written evidence (AIC0103).....	593
Dr Samantha Gallivan – Written evidence (AIC0185)	602
Ms Joanna Goodman, Dr Paresh Kathrani, Dr Steven Cranfield, Chrissie Lightfoot and Michael Butterworth – Written evidence (AIC0104).....	609
Google – Written evidence (AIC0225)	618
Government of Canada – Written evidence (AIC0222)	630
Government of China – Written evidence (AIC0145)	636
Government of Japan – Written evidence (AIC0224).....	637
Government of the Republic of Korea – Written evidence (AIC0228).....	640
Dr Paul Graham, Professor James Marshall, Professor Thomas Nowotny and Dr Andrew Philippides – Written evidence (AIC0088)	645
Guide Dogs – Written evidence (AIC0040).....	646
Dr Ozlem Gurses, Dr Antonios Kouroutakis, Dr Valentina Rita Scotti, Dr Aysegul Bugra and Matthew Channon – Written evidence (AIC0051)	653
Baroness Harding of Winscombe – Written evidence (AIC0072).....	654
HM Government – Written evidence (AIC0229)	660
Fabia Howard-Smith and Laurence Freeman – Written evidence (AIC0147)	672

The Human Rights, Big Data and Technology Project – Written evidence (AIC0196)	680
Dr Catrin Fflûr Huws – Written evidence (AIC0008)	690
IBM – Written evidence (AIC0160)	693
IEEE European Public Policy Initiative Working Group on ICT – Written evidence (AIC0106)	702
IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems – Written evidence (AIC0100).....	709
Imperial College London – Written evidence (AIC0214).....	710
Information Commissioner’s Office – Written evidence (AIC0132)	715
Information Systems Audit and Control Association (ISACA) London Chapter – Written evidence (AIC0193)	734
Information Technology Industry Council (ITI) – Written evidence (AIC0176) .	746
Innovate UK – Written evidence (AIC0220)	750
The Institute of Chartered Accountants in England and Wales – Written evidence (AIC0041)	768
Institute of Mathematics and its Applications – Written evidence (AIC0107) ...	776
International Associates – Written evidence (AIC0003)	783
Dr Maria Ioannidou – Written evidence (AIC0082)	786
Brian Joyce and Dr Ian Morgan – Written evidence (AIC0179)	795
Dr Paresh Kathrani, Dr Steven Cranfield, Chrissie Lightfoot, Michael Butterworth and Ms Joanna Goodman – Written evidence (AIC0104).....	796
Kemp Little LLP – Written evidence (AIC0133)	797
Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher and Professor Alan Bundy – Written evidence (AIC0029)	804
Dr Ben Kirman, Dr Conor Linehan, Dr Dan O'Hara and Professor Shaun Lawson – Written evidence (AIC0127)	805
The Knowledge, Skills and Experience Foundation – Written evidence (AIC0044)	806
Dr Ansgar Koene – Written evidence (AIC0208)	815
Dr Antonios Kouroutakis, Dr Valentina Rita Scotti, Dr Aysegul Bugra, Matthew Channon and Dr Ozlem Gurses – Written evidence (AIC0051).....	824
KPMG LLP – Written evidence (AIC0211)	825
Martina Kunz, Andrew Ware, Dr Simon Beard, Dr Seán Ó hÉigearthaigh and Dr Shahar Avin – Written evidence (AIC0150).....	835
Maciej Kuziemski and Toby Phillips – Written evidence (AIC0197)	836

Professor Marta Kwiatkowska – Written evidence (AIC0190)	837
Dale Lane – Written evidence (AIC0059)	839
Law and Innovation Research Group and the Legal Teaching Research Group from The Fundação Getúlio Vargas School of Law, São Paulo, Brazil – Written evidence (AIC0177)	843
Dr David Lawrence and Dr Sarah Morley – Written evidence (AIC0036)	848
The Law Society of England and Wales – Written evidence (AIC0152)	849
James Lawson – Written evidence (AIC0073)	857
Professor Shaun Lawson, Dr Ben Kirman, Dr Conor Linehan and Dr Dan O'Hara – Written evidence (AIC0127)	871
Professor Mark Lee – Written evidence (AIC0093)	872
Leverhulme Centre for the Future of Intelligence – Written evidence (AIC0182)	876
Leverhulme Centre for the Future of Intelligence – Supplementary written evidence (AIC0236)	886
Leverhulme Centre for the Future of Intelligence – Supplementary written evidence (AIC0238)	889
Leverhulme Centre for the Future of Intelligence and Centre for the Study of Existential Risk – Written evidence (AIC0237)	894
Leverhulme Centre for the Future of Intelligence and Centre for the Study of Existential Risk – Supplementary written evidence (AIC0239)	899
LexisNexis UK – Written evidence (AIC0164)	904
Liberty – Written evidence (AIC0181)	912
Chrissie Lightfoot, Michael Butterworth, Ms Joanna Goodman and Dr Paresh Kathrani and Dr Steven Cranfield – Written evidence (AIC0104)	926
Dr Conor Linehan and Dr Dan O'Hara, Professor Shaun Lawson and Dr Ben Kirman – Written evidence (AIC0127)	927
Professor Rosemary Luckin – Written evidence (AIC0246)	928
Dr Mike Lynch – Written Evidence (AIC0005)	930
Dr Mike Lynch – Supplementary written evidence (AIC0230)	936
Ms Nika Mahnič and Professor Kathleen Richardson – Written evidence (AIC0200)	942
The Market Research Society – Written evidence (AIC0130)	943
Professor James Marshall, Professor Thomas Nowotny, Dr Andrew Philippides and Dr Paul Graham – Written evidence (AIC0088)	947
Dr Neil McBride – Written evidence (AIC0047)	954
Mr John McNamara – Written evidence (AIC0081)	959

Sabine McNeill – Written evidence (AIC0009)	964
Professor Andrew McStay – Written evidence (AIC0015)	969
medConfidential – Written evidence (AIC0063)	982
medConfidential – Supplementary written evidence (AIC0244)	997
The Medicines and Healthcare products Regulatory Agency (MHRA) – Written evidence (AIC0134).....	1001
Microsoft – Written evidence (AIC0149)	1009
Dr. Zdenek Moravcik – Written evidence (AIC0019).....	1018
Dr Ian Morgan and Brian Joyce – Written evidence (AIC0179)	1021
Dr Sarah Morley and Dr David Lawrence – Written evidence (AIC0036)	1029
National Data Guardian for Health and Care – Written evidence (AIC0143)...	1037
Professor John Naughton – Written evidence (AIC0144).....	1042
NCC Group plc – Written evidence (AIC0240)	1046
Dr Jean-Christophe Nebel – Written evidence (AIC0102)	1056
Hadley Newman – Written evidence (AIC0155)	1058
Hadley Newman – Supplementary written evidence (AIC0156)	1069
Nominet – Written evidence (AIC0131)	1082
Norton Rose Fulbright LLP – Written evidence (AIC0079)	1086
Professor Thomas Nowotny, Dr Andrew Philippides, Dr Paul Graham and Professor James Marshall – Written evidence (AIC0088).....	1093
NVIDIA – Written evidence (AIC0212)	1094
Ocado Group plc – Written evidence (AIC0050).....	1108
Mr Jeremy O’Connor – Written evidence (AIC0034)	1115
Dr Dan O’Hara, Professor Shaun Lawson, Dr Ben Kirman and Dr Conor Linehan – Written evidence (AIC0127)	1119
Dr Seán Ó hÉigeartaigh, Dr Shahar Avin, Martina Kunz, Andrew Ware and Dr Simon Beard – Written evidence (AIC0150).....	1125
Dr James O’Shea – Written evidence (AIC0226).....	1126
Alex Olson – Written evidence (AIC0002)	1136
Onfido – Written evidence (AIC0163).....	1138
Online Dating Association – Written evidence (AIC0110)	1141
ORBIT The Observatory for Responsible Research and Innovation in ICT – Written evidence (AIC0109)	1145
Ordnance Survey – Written evidence (AIC0090).....	1150
Marion Oswald and Sheena Urwin – Written evidence (AIC0068).....	1156
Professor Maja Pantic – Written evidence (AIC0215).....	1168

Dr Andrew Pardoe – Written evidence (AIC0020)	1175
Joshua Parikh – Written evidence (AIC0031)	1179
Jonathan Penn – Written evidence (AIC0198)	1186
PHG Foundation – Written evidence (AIC0092)	1189
Dr Andrew Philippides, Dr Paul Graham and Professor James Marshall, Professor Thomas Nowotny – Written evidence (AIC0088).....	1197
Toby Phillips and Maciej Kuziemski – Written evidence (AIC0197)	1198
Professor Barbara Pierscionek and Dr John Rumbold – Written evidence (AIC0046)	1204
Professor John Preston – Written evidence (AIC0014).....	1209
PricewaterhouseCoopers LLP (PwC) – Written evidence (AIC0162)	1212
Privacy International – Written evidence (AIC0207).....	1223
Raymond Williams Foundation – Written evidence (AIC0122).....	1235
Professor Chris Reed – Written evidence (AIC0055).....	1242
Research Councils UK – Written evidence (AIC0142).....	1249
Research into Employment, Empowerment and Futures Centre (REEF), The Open University – Written evidence (AIC0124).....	1260
Professor Kathleen Richardson and Ms Nika Mahnič – Written evidence (AIC0200)	1270
Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy and Professor Simon King – Written evidence (AIC0029).....	1274
Mrs Violet Rook – Written evidence (AIC0151)	1275
Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy, Professor Simon King and Professor David Robertson – Written evidence (AIC0029).....	1277
Royal Academy of Engineering – Written evidence (AIC0140)	1278
The Royal College of Radiologists – Written evidence (AIC0146)	1294
The Royal Society – Written evidence (AIC0168).....	1304
The Royal Statistical Society – Written evidence (AIC0218)	1319
The RSA – Written evidence (AIC0157)	1327
Dr John Rumbold and Professor Barbara Pierscionek – Written evidence (AIC0046)	1336
SafeToNet – Written evidence (AIC0087)	1337
Sage – Written evidence (AIC0159).....	1347
Professor Maggi Savin-Baden and Eur. Ing. David Burden – Written evidence (AIC0061)	1361

SCAMPI Research Consortium, City, University of London – Written evidence (AIC0060)	1362
Dr Valentina Rita Scotti, Dr Aysegul Bugra, Matthew Channon, Dr Ozlem Gurses and Dr Antonios Kouroutakis – Written evidence (AIC0051).....	1369
Dr Huma Shah and Professor Kevin Warwick – Written evidence (AIC0066)..	1370
Professor Noel Sharkey – Written evidence (AIC0248)	1377
Simul Systems Ltd – Written evidence (AIC0016).....	1387
Jonathan Sinclair – Written evidence (AIC0023)	1391
Jonathan Sinclair – Supplementary written evidence (AIC0035)	1397
SiteFocus Incorporated – Written evidence (AIC0187)	1399
Dr Will Slocombe – Written evidence (AIC0056)	1404
Dr Chris Steed – Written evidence (AIC0017)	1409
Professor Richard Susskind – Written evidence (AIC0194)	1416
Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson and Dr Michael Rovatsos – Written evidence (AIC0029).....	1423
techUK – Written evidence (AIC0203).....	1424
Thames Valley Police – Written evidence (AIC0125)	1440
Thomson Reuters – Written evidence (AIC0223)	1445
Touch Surgery – Written evidence (AIC0070)	1447
Transport Systems Catapult – Written evidence (AIC0158).....	1449
Richard Tromans – Written evidence (AIC0227)	1456
UCL Knowledge Lab – Written evidence (AIC0105)	1461
UK Computing Research Committee – Written evidence (AIC0030).....	1471
The Association for UK Interactive Entertainment (Ukie) – Written evidence (AIC0116)	1478
Dr Ozlem Ulgen – Written evidence (AIC0112).....	1487
University College London (UCL) – Written evidence (AIC0135)	1494
Sheena Urwin and Marion Oswald – Written evidence (AIC0068).....	1506
Michael Veale – Written evidence (AIC0065).....	1507
Professor Chris Voss – Written evidence (AIC0118)	1511
Dr Toby Walsh – Written evidence (AIC0078)	1517
Andrew Ware, Dr Simon Beard, Dr Seán Ó hÉigearthaigh, Dr Shahar Avin, Martina Kunz, – Written evidence (AIC0150).....	1520
Professor Kevin Warwick and Dr Huma Shah– Written evidence (AIC0066)...	1521

Warwick Business School University of Warwick – Written evidence (AIC0117)	1522
Weightmans LLP – Written evidence (AIC0080)	1528
Wellcome Trust and the Association of Medical Research Charities (AMRC) – Written evidence (AIC0202)	1537
Vishal Wilde – Written Evidence (AIC0004)	1543
Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos and Professor Austin Tate – Written evidence (AIC0029)	1551
Professor Rebecca Williams – Written evidence (AIC0206)	1552
Professor Michael Wooldridge – Written evidence (AIC0174)	1558
Workday Inc. – Written evidence (AIC0183)	1564
Young Enterprise – Written evidence (AIC0091)	1568
Dr Jianhan Zhu – Written evidence (AIC0045)	1573
Diego Zuluaga – Written evidence (AIC0235)	1576

10x Future Technology – Written evidence (AIC0024)

Question 1:

What is the current state of artificial intelligence and what factors have contributed to this?

- i) By far the most widely-used AI technique in use in practical settings is “deep learning.” Deep learning is a term that describes an approach that uses large amounts of data to train “deep” neural networks (neural networks that have many layers.) Three factors have contributed to this development:
1. Organisations are collecting and storing unprecedentedly large amounts of data (“Big Data”)
 2. Processing power has grown enormously thanks to cloud computing
 3. Algorithms for multi-layer neural networks have reached maturity

How is it likely to develop over the next 5, 10 and 20 years?

- ii) In the next five-to-ten years, maturity is likely to come in the fields of sentiment mining (understanding how a person feels from the language that they use), behavioural prediction and autonomous vehicles. Beyond that, it is very difficult to predict, as there are likely to be many more enabling technologies (around wearables, nearables, or the internet of things, for example) that allow currently unforeseen developments.
- iii) The situation seems analogous to that in the early 1990s with the rapid development of internet technologies. Innovation is unevenly distributed, with smaller, disruptive organisations able to leapfrog established organisations with large technical infrastructures and cumbersome organisational structures.

What factors, technical or societal, will accelerate or hinder this development?

- iv) The greatest challenge that I foresee arises from the regulatory landscape in several industry sectors. For example, in the field of retail banking, an aversion to the risk (real or imagined) of running afoul of the Financial Conduct Authority (FCA) or Prudential Regulation Authority (PRA) slows the pace of innovation. The FCA is attempting to counter this with its Sandbox and Innovation Hub projects. However, the problem is circular: Until regulatory frameworks are overhauled to allow for AI, large organisations will be loath to innovate. But until innovators demonstrate what AI is capable of, regulatory frameworks cannot be updated. The development of the regulatory environment could either make or break the competitiveness of the UK as a centre for Fintech.

Question 2

Is the current level of excitement which surrounds artificial intelligence warranted?

v) The answer to that question depends on whose excitement we consider.

Warranted:

- Investment in AI technology for commercial purposes
- Consideration of legal and regulatory implications of AI in (*inter alia*)
 - medicine,
 - financial services,
 - education,
 - warfare,
 - eldercare
- Consideration of the effect on patterns of (un)employment
- Consideration of how bad actors may use AI

Unwarranted:

- Fears of immanent human-style general purpose intelligence
- Fears of AI autonomously turning against its makers

Question 3:

How can the general public best be prepared for more widespread use of artificial intelligence?

vi) There are several pressing social issues including: the potential end of collectivisation of risk with more personalised insurance modelling, the danger of using AI to infer sensitive data, people losing control of their personal data, and potential infringements to privacy. However all of these things can - with sufficient political will - be dealt with through democratic mechanisms of legislation and regulation.

vii) The largest upheaval is likely to arise from the profound economic impact of automation which, if left unaddressed, will lead to unsustainable income equalities, and social collapse. ¹

viii) The prospect of mass unemployment

Automation will not affect roles universally. Sectors likely to be hit include (but are not limited to)

- 1 transport,
- 2 manufacturing
- 3 retail,
- 4 farming,
- 5 scheduling, planning and management,
- 6 information management,

¹ Price Waterhouse Coopers "The economic impact of artificial intelligence on the UK economy" Jun 2017 <http://www.pwc.co.uk/economic-services/assets/ai-uk-report-v2.pdf>

7 back-office support functions

- ix) While the net positive impact for GDP is estimated at £200 billion in 2030, the gains will be unevenly distributed, with the majority of those gains accruing to individuals and organisations with the capital to invest in automation technology.
- x) The jobs created by AI are unlikely to be anywhere near numerous enough to offset the jobs that disappear. Some individuals, particularly those who are trained in engineering, creative and critical thinking, and interpersonal skills, will still be in demand. However, many others, in both low- and high-skilled roles, are likely to see some if not all of their functions made redundant.
- xi) In order to address both the widening disparity in income, and the erosion of the tax base, the UK government may be forced to consider both a universal basic income, and a taxation regime that taxes the economic outputs of automation as if they were income to human workers.²

Question 4: Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

- xii) Organisations that have the capital to invest in the development of intellectual property are the major beneficiaries of this trend towards the commercial use of AI. Those in the sorts of roles that are increasingly automated are the most obvious economic victims of this change. Societally, all citizens are at risk of an erosion of privacy and a reinforcement of the biases inherent in society.³

Question 5: Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

- xiii) There are analogies with the mid 1990s and the state of public understanding of the internet. Initially there was a great deal of confusion and hype as those in the media came to grips with what the internet actually was. As the associated technologies rolled out into workplaces and

² Malcolm James, "Could Bill Gates' plan to tax robots really lead to a brighter future for all?", *The Conversation* March 2017 "<https://theconversation.com/could-bill-gates-plan-to-tax-robots-really-lead-to-a-brighter-future-for-all-73395>

"South Korea introduces world's first 'robot tax'" *The Telegraph* August 2017 <http://www.telegraph.co.uk/technology/2017/08/09/south-korea-introduces-worlds-first-robot-tax/>

³ "Biased bots: Artificial-intelligence systems echo human prejudices" Princeton University April 2017 <https://www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices>

homes, the level of understanding grew and the public discourse naturally developed.

Question 6: What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

- xiv) All sectors stand to benefit, although these benefits will be unevenly distributed. Again, as with communication and the internet, the automation capabilities that AI affords will impact the economic and social landscape as a whole. In the short-to-medium term the gains will be most keenly experienced in tasks that:
- Can be routinised
 - Follow a pattern
 - Generate large amounts of data
- xv) Unlike previous industrial revolutions, this one will impact the service sector as much, if not more, than any other sector. Call centres are likely to all-but-disappear, much legal and medical work will be automated, and financial services will see major job losses.

Question 7: How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- xvi) This may require new legislative approaches to ensure that monopoly data power is regulated in the same way as other monopoly powers. Allowing individuals greater control of their own data - in particular, the right to opt out of data collection for certain purposes, or to request that data be deleted - is likely to be necessary in order to prevent abuses.
- xvii) It is also likely that, as privacy and security become more pressing concerns, informed consumers will shift towards services where there is a fair tradeoff between data collected and functions provided.
- xviii) Privacy can be considered as analogous to money or time: as a resource that we surrender in exchange for a desired outcome. Individuals try to get the best value for their money, or the greatest reward for their time, but inefficiencies, anticompetitive practices, and a lack of transparency often mean that they pay more for a good or service, or spend more time on inefficient service delivery, than they should. As a result, many industries are governed by regulatory bodies. The role of the ICO will need to evolve to understand data as a transactable resource, and to ensure that individuals get value for the data that they give.

Question 8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

xiv) There are several ethical impacts that require addressing:

1. As data becomes more valuable, the motivation to steal that data, or to otherwise obtain data without consent, will also increase. Deterrent penalties will need to rise concomitant with that increase.
2. As data increases in value as a resource for training artificial intelligences, there will be new criminal activities that involve data sabotage: either by destroying data, altering data, or injecting large quantities of misleading data. This is likely to require the creation of an entirely new class of prohibited activity (data sabotage) with its own penalties.
3. AI systems trained on data that reflect existing biases will reflect - and in some cases magnify - those biases.⁴ This poses a significant threat that socially marginalised groups will have even less access to employment, education, healthcare, financial services (among others) than is currently the case. It will be necessary to create meaningful deterrents to prevent organisations from negligently deploying decision-making systems that result in discrimination.
4. There is also a risk of highly-tailored propaganda impacting our democracy. While I must express some skepticism that Cambridge Analytica have managed to have as much of an impact on recent elections as their spokespeople claim, the aim of that organisation is clearly to target individual voters with persuasive messaging. This would create a worrying disparity between political organisations that have the financial and data resources to micro-target their election campaigns, and those that do not. The rules governing political campaigning must be brought up to date to reflect this.

Question 9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

xv) Black boxing should certainly be unacceptable in scenarios where decisions are made that impact people's lives and that could be subject to discrimination. That said, there are "model explainer" technologies that are becoming increasingly sophisticated - see LIME⁵ for example - that can be applied to models whose workings would previously have been inexplicable.

⁴"Machines taught by photos learn a biased view of women" Wired, 21 August 2017

<https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>

⁵ <https://arxiv.org/pdf/1602.04938.pdf>

xvi) Definition of what constitutes a “black box” cannot then be based on the learning algorithm used. Rather, it should be a functional definition that required that the features used to make a decision can be inspected *by the subject of that decision*.

Question 10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

No answer submitted

Question 11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

No answer submitted

24 August 2017

The Academy of Medical Sciences – Written evidence (AIC0210)

Summary

- The impact of artificial intelligence (AI) on biomedical research and the healthcare system is likely to be profound. Key benefits include improved efficiency of research and development processes, new methods of healthcare delivery, more informed clinical decision-making and empowerment of patients in managing their health. AI is already being used in these areas and this is certain to increase in the future.
- The ever-increasing volume of data being generated by the NHS, and through technology such as health apps and wearables, is further driving the development and use of AI.
- The key strength of AI is in rapidly analysing complex datasets. These data could be uninterpretable by a human or AI could automate existing human analyses, making interpretation faster and more accurate. It is expected that AI-based algorithms in healthcare will be used to complement the work of healthcare professionals but not fully replace them.
- The performance of AI is dependent on the quality of the data it uses. Therefore datasets should be high-quality and comprehensive to maximise the effectiveness of an AI algorithm and minimise the introduction of inaccuracies or bias.
- AI algorithms should be thoroughly tested and it should be shown that the system offers clinical benefit, accuracy and reliability over the alternative before implementation.
- Acceptability of AI and data sharing processes in healthcare should be informed by engagement with key stakeholders including patients and the public. Transparency around how and where AI is used is important to allow effective evaluation and validation of the system, and to enhance its trustworthiness amongst the public and key stakeholders.
- It is important to establish proportionate regulation of AI that balances appropriate safeguards against stimulation of innovation in this field.

Introduction

1. The Academy of Medical Sciences promotes advances in medical science and supports efforts to ensure that these are translated into healthcare benefits for society. Our elected Fellowship comprises some of the UK's foremost experts in medical science, drawn from a diverse range of research areas, from basic research through clinical application to commercialisation and healthcare delivery.
2. We welcome the opportunity to respond to this call for evidence on the implications of advances in AI. The Academy is monitoring the developments

in, and applications of, AI in medical science and healthcare through various workstreams: improving the health of the public in 2040; enhancing the use of scientific evidence; health apps; real world evidence; multi-morbidity; and regulation and governance of health research.^{6,7,8,9,10} This work has informed our input to relevant consultations such as the House of Commons' Science and Technology Committee's inquiry on algorithms in decision-making.¹¹

3. This response outlines some of the opportunities and challenges that use of AI may have for medical research and healthcare. The response is based on our recent policy work and the views of the Academy's Fellows and other experts with whom we collaborate.
4. AI refers to systems used to simulate human intelligence, and is a growing field due to increases in computational power that allow processing and analysis of large and complex datasets. Much of AI today exists in the form of machine learning, where algorithms use a set of training data to learn how to spot patterns in datasets that would otherwise be too complex for human analysis. It is expected that future developments will lead towards systems which interact with humans more directly, particularly when paired with robotics.

Implications for biomedical science and research and development

5. AI is being increasingly applied to further our understanding of basic science by detecting patterns or features that have been previously missed by researchers or are too complex for humans to identify. It is also used for a variety of functions across research and development (R&D) including computer-assisted drug design, clinical trial data interpretation and clinical trial simulations such as pharmacological modelling.
6. Randomised clinical trials (RCTs) are often used to generate information on the safety, efficacy and effectiveness of medicines. However, interventions are tested on a sub-section of a population group that meets eligibility criteria, for example age or number of conditions, which means that it can be challenging to generalise results to the wider 'real' patient population. Our Fellows have suggested that AI-simulated trials can make RCT results more

⁶ Academy of Medical Sciences (2016). *Improving the health of the public by 2040*. <https://acmedsci.ac.uk/file-download/41399-5807581429f81.pdf>

⁷ Academy of Medical Sciences (2016). *Real world evidence*. <https://acmedsci.ac.uk/file-download/38667-573d8796ceb99.pdf>

⁸ Academy of Medical Sciences (2015). *Health apps: regulation and quality control*. <https://acmedsci.ac.uk/file-download/37073-552cc937dcfb4.pdf>

⁹ Academy of Medical Sciences (2015). *Multiple morbidities as a global health challenge*. <https://acmedsci.ac.uk/file-download/38330-567965102e84a.pdf>

¹⁰ Academy of Medical Sciences (2016). *Regulation and governance of health research: five years on*. <https://acmedsci.ac.uk/more/events/regulation-and-governance-of-health-research-five-years-on>

¹¹ Academy of Medical Sciences (2017). *Response to the House of Commons Science and Technology Committee inquiry into algorithms in decision-making*. <https://acmedsci.ac.uk/file-download/79291192>

applicable to real world usage, and could also be used for license expansions (for example beyond the original population in which a drug was approved, such as in the elderly) or drug repurposing without the need for expensive and lengthy Phase III trials.

7. AI has the potential to utilise the increasingly large and complex pool of data collected through multiple sources such as wearable devices, health monitors and genome sequencing, with implications for both research and clinical care. As such datasets become more accessible, this opens up the possibility for greater patient and public involvement in (PPI) research, and the commercial sector is likely to be a major driver in this area with initiatives such as Google DeepMind Health and IBM Watson Health in development. In addition, the NHS offers a unique source of health data, presenting the opportunity for academic and commercial research to partner with the NHS in developing new AI tools. In such cases, it would be desirable for the research outputs to be developed and made available in collaboration with the NHS. Alongside

Examples of AI used in research and development

- An example of AI used in research is a Stanford-developed algorithm that, using histological images, uncovered new morphological features of breast cancer that hadn't previously been identified by clinicians using the same images.⁷
- DIYgenomics is a non-profit organisation that allows members of the public to contribute their health and genetic data for use in AI driven studies.⁸

research, smart-phone apps and wearable devices that monitor health measures such as heart rate or distance walked can be linked to GP surgeries to send data for use in clinical care.

Implications for the healthcare system and health outcomes

8. AI is becoming increasingly commonplace in healthcare, where it is routinely applied to calculate risk, aid diagnosis and generate medical images. These tools can guide the clinician, and others, through the diagnosis and decision-making process and support early intervention alongside prediction and prevention of future health problems.
9. Algorithms such as decision-support tools are key for supporting clinicians in making informed decisions about disease management, and can enable patients to take a more active role in decision-making. This is particularly important in choosing the best route of care in complex cases, such as circumstances where a number of medical conditions may need to be considered within the limited time available in a GP consultation. In addition, AI can enable automatic flagging of 'next steps' to a clinician when certain patient data is inputted, such as identifying the need to carry out specific diagnostic tests. However, the clinician-patient relationship should remain an

integral part of care.¹² There will remain situations where a clinician is best placed to optimise care based on clinical experience and context, and so AI should be used to complement clinical care but not replace the need for healthcare professionals.¹³

10. Clinical decision-support tools should be the subject of research evaluation and supported by funders. NICE, in discussion with NHS Choices, should coordinate the development of these tools based on the evidence generated by them.¹⁴
11. Increased use of AI will lead to changes in the skillset required for professionals, and training programmes should reflect this to allow staff to maximise on the opportunities afforded by AI. As such, there is a need to identify and address any gaps in capability to ensure the necessary training for the integration, manipulation and analysis of the data within appropriate ethical and regulatory frameworks.¹⁵
12. As with all innovations, there is a risk of inequity of access to the applications developed from AI and this should be a consideration for commissioners, particularly if the application has been developed using publically-generated data sets.

Example of AI in health and social care

An example of an emerging diagnosis aid is a University of Washington School of Medicine study that used 100,000 optical coherence tomography images to train an algorithm to detect age-related macular degeneration. The algorithm achieved sensitivities and specificities of over 90% and could therefore be used for automated screening of patients.¹³

Regulation and governance of AI

13. The MHRA has guidelines for the requirements of digital medical devices such as apps and implants and the laws that cover their use.¹⁷ However, these guidelines do not specify the process for the validation of algorithms, AI and devices and it is currently unclear how these devices fit with the regulatory framework or local infrastructure for implementation and evaluation of digital

¹² Chewning B, et al. (2012). *Patient preferences for shared decisions: A systematic review*. Patient Educ Couns **86**, 9–18.

¹³ Academy of Medical Sciences (2017). *Enhancing the use of scientific evidence to judge the potential benefits and harms of medicines*. <https://acmedsci.ac.uk/file-download/44970096>

¹⁴ *Ibid*.

¹⁵ Academy of Medical Sciences (2016). *Improving the health of the public by 2040*. <https://acmedsci.ac.uk/file-download/41399-5807581429f81.pdf>

¹⁶ Lee CS, et al. (2017). *Deep learning is effective for the classification of OCT images of normal versus*

Age-related Macular Degeneration. *Ophthalmology Retina* **124**, 1090-1095.

¹⁷ MHRA (2014). *Medical device stand-alone software including apps*.

www.gov.uk/government/uploads/system/uploads/attachment_data/file/564745/Software_flow_chart_Ed_1-02.pdf

devices such as the Paperless 2020 initiative, or through Academic Health Science Networks (AHSNs), as proposed by the Accelerated Access Review.^{18,19}

14. It is important to establish further proportionate regulatory processes around AI that maintain appropriate safeguards whilst also fostering a facilitative environment for innovation in this field. In addition, regulation should not impact the ability for companies to develop in-house AI systems that may be used for R&D but that do not directly affect health.

Transparency and limitations of AI

15. AI systems should be open to scrutiny to allow validation of effectiveness, evaluation of their potential risks and biases and to promote trust amongst users, recognising the need to consider IP protection for commercial developers.
16. It is essential that AI-based algorithms that impact health are thoroughly tested and found to be robust prior to use. This can be tested by establishing that the system offers advantage, accuracy and reliability over the alternative before being implemented. As AI systems often improve over time as new data becomes available, new versions or updates must also be tested to ensure that they are as robust as the previous system, as this robustness cannot be assumed. Manufacturers should inform regulators of the changes to software, and regulation should be able to accommodate such iterative changes. In addition, dialogue between software developers and regulators should occur early on and throughout the design process to ensure that the software fulfils regulatory requirements and to allow thorough and timely appraisal.
17. The limitations of AI should be recognised as it is dependent on the data used to develop it and so may incorporate any biases present in the data. Socio-economic differences in access to digital technologies can accentuate such biases by limiting the availability of data that is fully representative of the population. An example of bias arising from incomplete datasets is a study that compared care given to women with breast cancer across affluent and deprived areas. A lack of data from women in deprived areas missed the observation that they presented more advanced tumours than women from affluent areas.²⁰
18. Therefore testing and regulation should also include the propensity for algorithms to make errors and impart bias. These can be measured using test data and should be included in risk assessments. It is widely agreed that any

¹⁸ National Information Board and Department of Health (2014). *Personalised Health and Care 2020* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/384650/NIB_Report.pdf

¹⁹ Accelerated Access Review: Final Report (2016). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/565072/AAR_final.pdf

²⁰ Macleod U & Watt GCM. (2008). *The impact of consent on observational research: a comparison of outcomes from consenters and non consenters to an observational study*. BMC Medical Research Methodology **8**, 1-6.

algorithm used in clinical practice should undergo the same scrutiny as any new guideline or tool, including efficacy and risk analysis. Therefore there is a need for clear guidelines to assess acceptable risk and determine culpability in case an error is made or the performance of the algorithm falls below certain standards. This may require scrutiny of the methods employed by the algorithm.

Data sharing and privacy

19. The accuracy and robustness of algorithms is dependent on the quality of, and access to, both the data used to build and test the algorithm and the data inputted into the model. Therefore enabling access to comprehensive, high-quality data sources is key. Further to this, it is vital to note the importance of data quality, as well as quantity, to ensure high-quality data collection.
20. Communication and engagement with patients, clinicians and other key stakeholders is essential to help them to understand the value of health data and how it is used by AI in research and healthcare. This can help them to make informed decisions about contributing and sharing data. Initiatives to increase public dialogue and understanding around this should be promoted and the Academy is pleased to be working with Understanding Patient Data on a piece of public dialogue to inform this area.²¹ Sharing of data, particularly with commercial bodies, can be contentious and there needs to be clarity and transparency around where, how and why data is shared for this purpose, with public acceptability being an important consideration.
21. In circumstances where publically generated data is shared for commercial use, it should be done so for the potential benefit to the health system or the public. Shared ownership of data between the NHS and commercial partners, or the IP generated from this data, could help to ensure that the exchange of data is of such benefit.
22. It is important to acknowledge that no mechanism of data anonymisation – particularly pseudo-anonymisation – will be entirely risk-free, but steps can be taken to minimise these risks. Appropriate safeguards which promote accountability and best practice in use of data, and appropriate sanctions for breaching data privacy, will help to reduce risks. In addition, good data governance practices are essential and these are supported by various guidance and legislation including the Information Commissioner's Office, the Government's response to the National Data Guardian's Review of data security, consent and opt-outs, and the new EU General Data Protection Regulation 2016, which comes into UK law in May 2018.²² The risk of manipulation of data or an algorithm by outside interference needs to be considered and appropriate safeguards and sanctions put in place to minimise the risk of such an event.

²¹ <https://understandingpatientdata.org.uk/>

²² [Department of Health \(2017\). *Your Data: Better Security, Better Choice, Better Care*.
www.gov.uk/government/uploads/system/uploads/attachment_data/file/627493/Your_data_better_security_better_choice_better_care_government_response.pdf](http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/627493/Your_data_better_security_better_choice_better_care_government_response.pdf)

23. Historically, patients give their consent for any aspect of their health data to be shared for a specific use. If the terms of use change, re-consent is usually required to ensure the patient remains informed about the use of their data. The Government's response to the recommendations of the National Data Guardian's recent Review accepts the proposed changes to this model in favour of a system more centred on 'opt-outs'.²³ Consent models across the health system should be homogeneous and standardised to ensure that patients are informed and developers understand what data is available to them.

This response was prepared by James Squires (Policy Officer) and Luiz Guidi (Policy Intern) and was informed through the Academy's previous activities and consultation.

Academy of Medical Sciences

11 September 2017

²³ *Ibid.*

Accenture UK Limited – Written evidence (AIC0191)

**RESPONSE TO “SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE –
CALL FOR EVIDENCE”
ACCENTURE (UK) LIMITED, 6 SEPTEMBER 2017**

Accenture’s definition of Artificial Intelligence (“AI”)

- 1. AI can be defined as a constellation of technologies that allow smart machines to extend human capabilities and intelligence by sensing, comprehending, acting and learning—thereby allowing people to achieve much more than can be achieved without the technology.

The pace of technological change

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- 2. The Committee rightly asks what is the current state of AI, because there is often a high degree of hype in today’s public discourse. We see discussions ranging from AI becoming spontaneously evil weaponry to AI solving all the world’s problems. In many respects, the current state of AI is that we are not “there” yet; AI needs human helpers to execute its tasks with excellence. AI can undertake tasks but not necessarily end to end processes of any complexity. However, even in relatively “weak AI” there are many things that AI can do very well, especially where tasks are (1) repeatable and/or (2) require time-consuming analysis of large datasets.
- 3. The progress of AI has accelerated in recent years due to the rise of big data, the proliferation of connected sensors and actuators (the Internet of Things), and access to vastly increased and cheap processing power; especially through the cloud. AI is now becoming a commercial reality.
- 4. We are seeing that the technology is progressing rapidly, and we believe that the entirety of the technology market will be impacted by the emergence of AI. The AI technology market is likely to be highly valuable by the year 2020. While it is somewhat difficult to approximate where AI will be in 20 years, we do anticipate that in 2-3 years all IT services either will incorporate AI capabilities or will be at serious risk of obsolescence.
- 5. One key factor that will, in our view, play a significant role in promoting the adoption of the technology will be the ability of human stakeholders - consumers, employees and citizens, among others - to trust the technology. That will rely on principles of honesty, fairness and transparency. This will also require policy interventions which address the question of liability in the event of misadventure, unintended consequence or bias.

- 6. See further:
https://www.youtube.com/channel/UCD8h3oTfr1CbanGIWnzIP_g/playlists?sort=dd&shelf_id=10&view=50

.....

The pace of technological change
Is the current level of excitement which surrounds artificial intelligence warranted?

.....

- 7. In our view, the enthusiasm and attention which AI is drawing is warranted from a long-term perspective, with an understanding that the next 3-5 years are critical years of exploration, planning and investment. Now is the opportunity for the UK to capitalise on the opportunity that AI presents for the growth of the economy and our society.
- 8. Our research findings suggest that the development and roll out of AI will lead to a significant economic boost and additional gross value added (GVA) to the global economy. Businesses successfully applying AI could increase profitability by an average of 38 percent by 2035. However, to thrive (and to continue to exist) in 2035, businesses must begin with smart planning and investment today in three key areas: (1) technology; (2) data; and (3) people.
- 9. Equally AI technologies can help tackle some major societal challenges and create a more inclusive society, including pre-empting future skills gaps; supporting workers at risk of displacement through career transitions and reskilling; making the workplace more accessible for people with disabilities; providing upskilling, employment and entrepreneurship opportunities for groups lacking access e.g. young entrepreneurs. We are delivering on this through several 'live' projects in association with our partner ecosystem:
 - Project Drishti using Microsoft AI technology to improve workplace accessibility for blind or partially sighted Indians in collaboration with the National Association for the Blind India and Nascom.
 - Youth Business USA and Accenture have co-created an entrepreneurship platform called SkysTheLimit, which itself uses a mentor matching algorithm to connect young entrepreneurs with mentors.
 - Accenture ran an AI hackathon across 25 Accenture geographies – where Accenture teams competed to create solutions, using AI that could support students, job seekers and entrepreneurs build the skills to thrive in the digital economy.
- 10. For further insights, please refer to our reports, "Why Artificial Intelligence is the Future of Growth" and "Boost Your AIQ."

.....

Impact on society
How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

11. To prepare the public for widespread use of AI, UK industries and the UK Government need to create an environment in which AI can be trusted. While there are many issues that the widespread use of AI will raise, we highlight the following issues due to their impact on the ability to establish trust:
 - **Automated decision making** – automation of important decisions around personal issues such as mortgage applications, housing, employment and bank loans;
 - **Job losses** – the fear that AI will impact employment levels in an extremely significant way (the so-called ‘job apocalypse’) and will accelerate, not lessen, the gig economy, thereby eroding the meaningfulness of work and gainful employment;
 - **Authentication** – concerns around what is authentic in a world of AI, given that AI can be used to create seemingly accurate but fake news, identities, transactions and even memories; and
 - **Inclusion & Diversity (I&D)** – concerns around AI algorithms being developed by a narrow and homogeneous sector of the population leading to potential AI bias as well as accessibility of AI skillsets to under-served and/or underrepresented populations.
12. For automated decision making, the key will be to establish strong governance that embeds accountability into AI systems. This in turn requires an assessment of the types of decisions it is proposed be undertaken by AI that would require explanation or create an expectation of explanation. This would include, by way of example, in the areas of employment, recruitment, lending, education, healthcare, housing and safety. For more information, please visit our blog, [“Why explainable AI must be central to responsible AI”](#).
13. To work in AI, people will need an entirely new skillset. Companies must make radical changes to their training, performance and talent-acquisition strategies. In collaboration with other stakeholders, such as unions, government, educators, they will need to reskill quickly and retrain to, ensure current workers continue to be relevant and continuously adaptable. The Government must also equip citizens with the multidisciplinary and STEAM skills—science, technology, engineering, arts, and mathematics—required by the development of AI.
14. Workers should aim to not only be more productive, but to deliver more creative, precise and valuable work. This will involve fostering a culture of lifelong learning, much of it enabled by technology: personalized online courses that replace traditional classroom curricula and wearable applications such as smart glasses that improve workers’ knowledge and skills as they carry out their daily work.

15. Further, in order to address I&D, both the public and private sectors must make every effort to train those who were left behind by such fast-moving technological developments in the past: minorities, women, working mothers, disabled persons. Our educational approaches to AI must create an inclusive rotation to these new technologies such that traditionally under-served and underrepresented communities have equal access and opportunity.

.....
Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?
.....

16. There are some who raise concern that the benefits arising out of the development and use of AI and data will be held by the few (be that “the elites” or the “powerful 1%” of entities and individuals) and that those individuals or entities will benefit the most because they are best placed through the advantage inherent in their position to capitalise on the technology early, potentially locking out others from benefitting in the same manner. For the purposes of helping the general public to be better prepared for AI, it is right for governmental entities to consider how intervention and collaboration with business communities can mitigate potential disparities in terms of access and opportunity.

17. In terms of who is gaining the least, it is very possible that those communities who are under-served by technology are not only those who have less access and/or skills in terms of the technology, but for whom there is perhaps the greatest imbalance in that the types of data that are being collected are about them. That imbalance may be reflected in data which is potentially less helpful (be it in relation to access to finance, data related to obtaining new jobs or educational opportunity or data that does exist due to policing, welfare benefits, etc).

18. A Responsible AI philosophy, with its goal of providing better outcomes for all people, is our response to mitigating the potential disparities that some fear the development and use of AI could bring. To meet the Responsible AI imperative and take a people first approach, a collaborative effort between the government, business and broader society should:

- **Emphasise education and training**, especially for people who may be disproportionately affected;
- **Reinvigorate codes of ethics** by adapting for the many ways AI will impact how an organisation will operate and how its people will interact with each other and with AI;
- **Help create adaptive, self-improving regulation and standards** to keep pace with technological change;

- **Establish sound cybersecurity practices;**
- **Lower the barrier to entry for small business entrepreneurs -** actions include continuing to make governmental data sets available to the public in a low-cost, accessible and digestible way and governments encouraging the private sector and scientific and research institutions, to share data and collaborate over such platforms, which can help support the development of vibrant AI eco-systems. Governments can also remove regulatory obstacles to the analysis and testing of big data. Data mining is one area in which regulatory barriers remain, and which is key to machine learning.
- **Integrate human intelligence with machine intelligence** by reconstructing work to take advantage of the respective strengths of each.

.....
Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?
.....

19. Yes, we believe that efforts should be made to improve understanding of, and engagement with, AI. In the last year, there has been an ever-increasing velocity of articles, blogs, speeches and thinking raising concerns about AI. The reality is that, in some instances, AI itself does not introduce novel questions or concerns (for example, the classic *trolley* question that is being raised in the context of autonomous vehicles). Yet, the hype around the technology has the potential to create misguided understandings of both the promise and peril of AI. The focus on the vivid risks rather than the potential benefit is negative for the public overall as it potentially impedes adoption of the technology and thereby delays the uptake of the benefits brought by the technology.
20. Given the public’s increased interest in this area, this is the right time for educational campaigns around: (1) topics like data privacy and cybersecurity; and (2) helping consumers and citizens understand their rights in a world of AI. Local communities should consider hosting technology fairs to galvanize their citizens around STEM education initiatives and to create further interaction and exposure to AI. This level of tangible engagement will enable an informed dialogue within our communities, accountability between public and private sectors as well as a drive to innovate.

.....
Industry
What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.
.....

21.Data in Accenture’s report on “How AI boosts industry profits and innovation” indicates that the manufacturing, professional services and retail sectors will benefit the most from AI adoption in terms of industry output. Other sectors set to benefit include public services, information and communication, financial services, healthcare and education. Yet, due to the possibilities of AI and its application, any business (regardless of its sector) which utilises the benefits of AI will be able to meet its customers’ demands in a more personalised way, and tailor its output to meet those demands. This will allow those businesses to capitalise upon their position in the market and obtain a competitive advantage.

22.The main consideration for any business is how it harnesses and governs the use of AI in a sustainable and ethical way. This is particularly critical given the grey areas in the regulatory space which will prevail longer than the innovation cycle. The key point is that any business which uses AI to disrupt in a responsible and a sustainable way, will stand to benefit the most in the future.

.....
Industry

How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

.....
23. As discussed more fully below, it is important to understand two key points in the context of Big Data. First, the concept of “Big Data” encapsulates both the volume of data and the ability to extract meaning from data, and thus, datasets have little intrinsic value absent the ability to access and analyse them. Without tools to extract meaning from them, datasets, whether small or large, are neither beneficial nor harmful. Second, there is widespread agreement that the analysis and use of Big Data can convey huge benefits, e.g. improve the quality of health care by reducing errors, save energy by controlling home electricity use and help automakers improve safety.

24.As innovation depends on data collection and analysis, any regulation targeting the use of big data should take this into account. We need to enable innovation that feeds the benefits of a data-driven future whilst ensuring that it contributes to and safeguards the public good. For example, Citymapper relies on transport data from the UK Government and Greater London Authority to allow users to choose a method of transport and travel across London based on information such as price, delays and weather.

25.Further, governments have a key role to play in this respect, particularly in opening-up data to small enterprises—which unlike large corporations might not have the resources to accumulate a critical mass of data with which to innovate. The most critical levers to help small enterprises take advantage of AI are access to data, technology and people. The UK Government can lead by example by sharing public-sector data-sets through the creation of public-data platforms that small enterprises can freely access. In addition, it can

- encourage the private sector and scientific and research institutions to share data and collaborate over such platforms, which can help support the development of vibrant AI eco-systems, as mentioned above.
26. To help ensure that data is safeguarded, in the context of the security of AI, we support the principle of Security by Design. Solid cybersecurity practices are an essential underpinning to ethical design to ensure that consumer and citizen trust is established and maintained throughout the entirety of an AI deployment cycle. Industry and governments should work together at a global level to create a common understanding of existing international security standards, certifications and methodologies that support the security of Information, Communications and Technology (ICT) products and services and which can be relevant for AI applications.
27. In our view, the notion of acquiring large datasets to abuse and protect a dominant position will be difficult in practice. First, the argument fails to account for the ability of more than one company to access and use the same datapoints, e.g. a consumer may use more than one online navigation service where both can collect time, travel and location information. Second, platforms that collect large datasets have a difficult time preventing competitors from creating a similar dataset or datasets, regardless of size or similarity, that serve the same outcome or functionality. Third, it is difficult to disrupt a competitive consumer market merely by acquiring large datasets. As stated above, datasets have little intrinsic value without the ability to extract meaning from them. The dataset is a necessary-but not the only component of delivering meaningful insights from data. Having the tools to analyse it and the experience to understand its meaning are the others. The correct question to ask therefore is not whether it is difficult to enter and compete in a market because of high entry barriers, but rather whether the owner of a dataset has escaped the bounds of a competitive marketplace and illegally exploited that advantage by manipulating consumers into accepting higher prices or lower quality goods. On that, regulators must carefully evaluate the potential harms against the benefits (including entry of new products and services into the marketplace) which the use of large datasets confers.
28. To the extent regulators feel it necessary to investigate an abuse of a dominant position in the digital marketplace, existing UK competition laws are already well-suited to protect consumers. It is not necessary to look to privacy laws or other legal disciplines to correct abuses by undertakings with market power in the area of data collection and analysis. Conversely, just as existing UK competition laws are well-suited to address competitive harm related to data collection or use, existing UK privacy regulations can already adequately address any potential privacy-related concerns as to the possession or use of big datasets.

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this

question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

29. In our experience, the sorts of questions and concerns raised by stakeholders around AI include the following:
- Job apocalypse - AI is so ruthlessly efficient that it will lead to massive job losses.
 - The singularity- We will create something that is more intelligent than humans and we will lose control.
 - Inclusion & diversity - How do we eliminate bias arising out of the technology? How do we avoid the impact of the technology most adversely impacting those who we have struggled to help be socially mobile?
 - Data privacy - AI will erode our notions of data privacy and do things with data that we didn't consent to.
 - Lack of transparency - AI doesn't explain itself.
 - Artificial stupidity - AI is actually not sufficiently intelligent right now, and this could lead to discriminatory actions.
 - Stewardship of data - AI has access to large datasets, how can we be good stewards of our customers' data.
 - Authentication - How will I know when I am dealing with something that is not true, real or human?
30. The involvement of ethics and core values are seen as an antidote to these concerns. What is required is a way, in practice, to apply ethical values to the questions and concerns described above.
31. Having an appropriate framework addresses this need. At the heart of such a "Responsible AI" framework, is governance, ethical design and deployment strategies, monitoring and auditing of the outputs of systems, together with structures that allow sufficient transparency.
32. This framework should include global, industry driven guidelines, standards and best practices to develop trust for the AI-driven systems and business models and permit the flexibility for innovation, allowing codes to develop with the technology. Current examples of these guidelines include, the IEEE Global Initiative for Ethical Considerations in AI and AS and its nine pipeline standards on Ethically Aligned Design.
33. Humans are central to any framework developed to combat the potential negative ethical implications of AI. This consideration not only concerns the structure and systems of the framework but also the relevant human stakeholders. For example, if organisations use AI to undertake roles previously undertaken by employees, then they need to consider how they can re-skill the affected employees and the local communities.

- 34. Ultimately, we need to appreciate that since there may be bias in any systems using AI, the focus should be to seek to minimise the impact of any bias by creating governance structures that are transparent enough to enable us to detect and correct flaws in the systems.
- 35. For further information, please refer to: <https://www.linkedin.com/pulse/can-we-trust-artificial-intelligence-answer-ai-christina-demetriades> and http://standards.ieee.org/develop/indconn/ec/rfi_responses_document.pdf

.....

In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

.....

- 36. Firstly, we need to consider that AI systems are developed in a human centric environment (where unfortunately a degree of bias needs to be regarded as inevitable) and then consider when we can deem it appropriate to use an AI "black box" system.
- 37. On occasions, there may not be absolute transparency in AI – AI systems may not always be able to be developed with the complete end-to-end understanding of how a particular type of AI works, and how to trace and audit decisions made by AI. In some circumstances, complete transparency may be unachievable and in others it is possible it could even be undesirable. However, the material issues in ethical AI are honesty and fairness and these ends are not necessarily delivered automatically by decision processes that are absolutely transparent.
- 38. Second, consumers and citizens will not permit a lack of transparency in those areas that are most personal to them, for example, decisions around education, housing, mortgage financing or employment. The key here is how organisations can enable their systems such that there is explainable AI.²⁴ While the technology may not necessarily fully enable this today, organisations must anticipate that consumers want sufficient explanations and that organisations will be accountable for their use of the AI. Regardless of the GDPR's position on this topic, it is inevitable that consumers are shifting from an expectation of privacy to an expectation of explanation. For example, organisations should provide a clear ability for the recipients of AI driven services to appeal errors or questionable decisions developed by AI. In this way, the use of AI systems is viewed through the lens of the broader social and ethical context.
- 39. Please refer to: http://standards.ieee.org/develop/indconn/ec/rfi_responses_document.pdf#blank and <https://www.accenture.com/us-en/blogs/blogs-why-explainable-ai-must-central-responsible-ai?src=SOMS>

.....

The role of the Government

²⁴ The United States Defense Advanced Research Projects Agency defines Explainable AI as AI systems that have the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future.

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- 40. The UK Government should take an active role in developing and using AI in order to keep pace with the global digital economy.
- 41. The UK has already allocated £17m to AI research and should continue to take an active role in investment in AI research and facilitating a vibrant ecosystem around R&D hubs. The participants in any major technological movement—whether they are a government, corporation, entrepreneur, inventor, or civil society—need a safe space to go to share ideas, develop best practices and solve problems. But perhaps most critically, to allow technology transfer from basic research to applied research. For example, hubs such as those in the universities of Cambridge and Oxford, have stimulated the creation and development of several start-ups that achieved major AI breakthroughs and later became prime acquisition targets.
- 42. In addition to supporting investments in innovation, the Government should look for ways to incorporate AI wherever possible as early adopters. National governments should ensure their own datasets can be leveraged by AI stakeholders by creating new public data platforms that enable open access. Governments should encourage the setup of democratized databases by private sector players for sharing anonymized and encrypted data that can be publicly accessed to support development. For example, Australia’s Department of Human Services is applying Microsoft AI technology to help employees respond faster to citizens’ inquiries.
- 43. Our view is that Government has a role to work with industry and other AI stakeholders to develop a smart regulatory framework that addresses the issues that arise in the use and application of AI (rather than the technology *per se*), including data protection, ethics and transparency, IP ownership, risk allocation and cyber-security. This framework should include global, industry driven guidelines, standards and best practices that can help create and safeguard trust at the heart of AI-driven systems and business models and permit the flexibility for innovation, allowing codes to develop with the technology. Examples include, the IEEE Global Initiative for Ethical Considerations in AI and AS and its nine pipeline standards on Ethically Aligned Design.
- 44. These multi-stakeholder initiatives have the strongest influence on creating industry cross-fertilization and equal access on AI resources for entrepreneurs and big players alike. They also help to identify gaps in existing standards and certifications, which ecosystem players can then act upon. A one-size fits-all solution should be avoided as applications vary across sectors and industries. See further: <https://www.accenture.com/gb-en/insight-engage-digital>.

Learning from others

What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

45. A number of governments around the world have recognized AI as a key driver of future growth and have recently announced strategies to substantially increase investment in AI. Examples include:
- **EU:** Robotics PPP (SPARC): €700 million allocated to research. SPARC is believed to be the biggest civilian research programme in this area in the world and aims to develop a robotics strategy for the region.
 - **China:** New Development Plan announced on 17 July to create \$150bn domestic AI industry by 2030. This includes promoting interdisciplinary research to connect AI with other areas and funding moon shoot projects.
 - **Canada:** Pan-Canadian Artificial Intelligence Strategy, supported by \$125million for AI research.
 - **Singapore:** National AI Programme investing \$150m over next 5 years, focused on growing knowledge and developing talent in AI.
 - **France:** Announced its #FranceIA strategy in January, with significant investments planned particularly for start-ups.
 - **Japan:** In addition to making investments in R&D, Japan, considered a fast follower in AI, has set out a number of actions to build its AI capabilities and create an environment for innovation, in its 2017 'Investments for the Future Strategy'. Actions include:
 - Creating regulatory sandboxes,
 - Promoting open data
 - Creating an ecosystem for start-ups
 - Strengthening industry-academia collaboration.
46. The UK has the opportunity to build on its AI capabilities and centres of excellence to become a leader in AI. The UK should consider developing a dedicated strategy supporting investment in AI, underpinned by significant public and private financial investment and an environment that supports innovation.

Accenture (UK) Limited

6 September 2017

Advanced Marine Innovation Technology Subsea Ltd – Written evidence (AIC0038)

For the attention of:- **Select Committee on Artificial Intelligence**

Call for evidence.

Answers submitted by:-

Eur.Eng. Ramsey Quayle Martin MBE BSc CEng FIMarEst
Director;
Advanced Marine Innovation Technology Subsea Ltd.

May I apologise in advance for statements that are contrary to what many people understand of computer systems.

First may I comment on the definition of “Artificial Intelligence”.

In respect of the digital computer I consider the use of the word “Intelligence” to be in contravention of the real meaning of the word.

A computer can only do precisely what it has been programmed to do. It will do exactly what has been programmed, right or wrong, with all programming errors rigorously included.

A digital computer can only recognise two states, namely 1 and zero. Digital programming merely creates strings of 1s and zeros which represent numbers and letters.

Therefore the “artificial intelligence” of the computer is simply no more than the mindless performance of what it has already been programmed to do.

Q1. Fundamentally the computer has not advanced in function since its original conception. All that has changed is the capability to perform the programme actions at higher speeds.

When I was first introduced to programming at the end of the 60s the message clearly stated was

“Do not consider the computer to be clever. In reality the computer is incurably stupid and will only mindlessly do exactly what you have programmed it to do.”

Apart from an increase in speed, and increase in volume of data that can be stored and a reduction in simple physical size of the machine nothing of a fundamental nature relating to digital calculation has changed since the 60s.

Q2. The current level of excitement surrounding “artificial intelligence” is not warranted. Most of it amounts to a response to the marketing hype by those with most to gain from controlling information and selling supposedly intelligent systems.

Q3. The best preparation for more widespread use of so called “artificial intelligence” is to limit divulging information to that which one can lose without risk.

Q4. Those gaining most are those who are producing the computer based systems and using the data that has been gathered.

Q5. The only way to improve the public understanding of so called “artificial intelligence” is to describe the reality in cold blooded un-emotive factual language. They must understand that their data can be readily stored and recovered.

Q6. The key sectors standing to benefit from the development and use are those who are writing the code and those using the code to control the flow of information. The other sectors will simply be squeezed by the controllers.

Q7. The data monopolies and “winner takes all” cannot be adequately addressed. That is a simple fact of life. The robber barons of the dark ages are still with us in a different guise.

Q8. It is highly unlikely that the ethical considerations and negative implications can be adequately resolved unless there are very severe sanctions and punishments for those who do not behave in a responsible and ethical manner. But in this world the unethical tend to have more “rights” than the ordinary people.

Q9. With regards to the issues of privacy, consent, safety, diversity and the impact on democracy the best way to understand the implications of what can be done with stored data is apply an open intelligent mind and read between the lines written by Orwell . i.e. 1984 and Animal Farm.

Q10. The only role that the government can take is to enforce privacy in respect of personal data.

Q11. There is little to be learned from other countries and organisations unless one stands back to take a constructive overview of the lemmings in a headlong rush for the cliff edge. All are heading down the same track.

30 August 2017

Agents, Interaction and Complexity (AIC) group, University of Southampton – Written evidence (AIC0115)

1. This response to your call for evidence has been compiled by Professor Timothy Norman, head of the Agents, Interaction and Complexity (AIC) group within the Electronics and Computer Science Department of the University of Southampton. It reflects a range of expert viewpoints from the group as a whole. We focus on two questions in the call: public perception (question 5); and industry (question 6).

Public perception

2. The hype around artificial intelligence and machine learning has led to many people holding false beliefs about the capabilities of AI. This includes inflated beliefs about both the positives and negatives of AI. Through our research involving deployments of smart energy systems based on AI in homes across the UK, for example, qualitative results demonstrate clearly that people over-estimate the capabilities of smart thermostats and sensors. In the same vein, many now believe killer robots will soon come about. These misinterpretations of the capabilities of AI builds a lot on science fiction rather than fact and it is important to educate the population in how such systems are built in order for them to understand what to expect from them. To this end, building on initiatives that aim to teach programming or to support a “maker culture”, we should introduce the basics of AI through similar programmes and also improve the population’s understandings of the mathematics behind AI. Television programmes as well as online media could be vehicles for such programmes.

Industry

3. There is a broad range of sectors that stand to benefit from the development and use of artificial intelligence. In our response we focus on two specific sectors that are prominent in current social and political concerns: transport and energy; and policing and security.
4. In transport, there is considerable debate and discussion about the future prospect of autonomous vehicles. More immediately, however, AI may play a key role in the drive towards the electrification of transport. Significant increases in the use of plug-in electric vehicles will place a considerable strain on local electricity networks, as well as increased demand on renewable supply. Through the use of energy storage, whether this uses stand-alone batteries or in an electric vehicle, it is possible to make more effective use of volatile renewable supplies. AI can be used to

manage the complexities of charging and discharging batteries in large-scale distributed systems, where every household can be a prosumer (e.g. producing energy from solar panels installed on their roofs), and parking stations can become virtual power plants. AI can be used for predicting supply and demand, but may also play an important role for autonomous and localised control through integration with the smart grid and the Internet of Things.

5. To give a specific example, we have explored how users can specify details of travel, including when they will need to use their car, and how far they are planning to drive. Automated auction mechanisms are then used to price their usage, or offer a revised price if, for example, they can be flexible in departure time. Here, people can save money if they can be more flexible with their charging allocation, which can take pressure away from the grid by charging vehicles in a more logical 'order'. Further details can be found at: <https://www.southampton.ac.uk/news/2016/07/green-electrics.page>
6. In security and policing, there are an increasing number of data analytics tools for mining sources such as social media. These are built upon machine learning algorithms, and hence AI methodologies. Outputs from these algorithms must, however, be interpreted by human experts, and, as noted in the call for evidence, it is not necessarily clear how this 'black box' output (question 9) has been generated. The problem is that human experts must use evidence such as this to make recommendations for interventions in a wide range of critical situations including potential terrorist attacks and paedophile offences. Through the use of structured reasoning models, developed from studies of how humans reason, software tools can support users to structure their interpretation of evidence from various sources. AI algorithms can then be employed to inform the human decision maker what the possible interpretations are: the plausible hypotheses, given the evidence as structured by the human analyst. The CISpaces system is one example of such a system. Further details can be found at: <http://www.cispaces.org>
7. An important message that underpins these examples is that artificial intelligence technologies are best (and, we would argue, only effectively) employed in conjunction with, and in support of human decision making. Neither the hype that argues that AI can solve all our problems nor that which argues it will replace human intelligence and insight are worthy of consideration.

6 September 2017

AGISI.org – Written evidence (AIC0184)

From: AGISI.org

Dr. Colin W. P. Lewis, A.I. Research Scientist

Prof. Dr. Dagmar Monett, A.I. Research Scientist (AGISI & Berlin School of Economics and Law)

1. a) **What is the current state of artificial intelligence?** There are currently no 'true' Artificial Intelligence (A.I.) systems. There are ad-hoc 'learning' systems, let's call them narrow A.I. systems.

Defining A.I. The literature abounds with definitions of A.I. and human intelligence although very little consensus has been reached to date. Our comprehensive research of A.I. practitioners worldwide, *Research Survey: Defining (machine) Intelligence* (Lewis & Monett, 2017), which has collected over 400 responses, has identified considerable interest in identifying a well defined definition and goal of A.I. We hope that the results of our survey help to overcome a fundamental flaw: "That artificial intelligence lacks a stable, consensus definition or instantiation complicates efforts to develop an appropriate policy infrastructure" (Calo, 2017).

The goal of A.I., closely linked to its definition and highlighted in our survey, should ensure the 'why' of Artificial Intelligence; however, very few research papers provide a robust goal with society-in-the-loop. We agree with Hutter (2005): "The goal of A.I. systems should be to be useful to humans." Or as Norbert Wiener wrote in 1960, "We had better be quite sure that the purpose put into the machine is the purpose which we really desire" (Wiener, 1960).

Whilst there are breakthroughs in narrow A.I. systems that can 'simulate' and surpass certain 'individual' aspects of human intelligence (for example, specific elements of pattern recognition, quicker at search, calculations, data analysis, and other cognitive attributes), A.I. development is currently some way off from achieving the goal of fully replicating human intelligence. However, the narrow A.I. methods, which are more specifically fields of A.I. research, are making considerable progress as stand alone techniques, namely Machine Learning (ML) and classes of ML algorithms such as Deep Learning (DL), Reinforcement Learning (RL), and Deep Reinforcement Learning (DRL).

Researchers acknowledge that the methodology applied in narrow A.I. systems can be unstable (Mnih et al., 2015). Nevertheless, these A.I. sub-domains are already starting to have considerable economic and social effect, as we outline below, and this impact will accelerate in the near future. Briefly:

- **Machine Learning:** Whereas the vast majority of computer programs are hand-coded by humans, Machine Learning algorithms are capable of 'self-

learning,' improving computability on a specific task against key performance metrics, and enhance output through experience.

- **Deep Learning:** The key aspect of deep learning is that its features are not designed by human engineers. Instead, "they are learned from data using a general-purpose learning procedure" (LeCun, Bengio & Hinton, 2015). Deep Learning is defined by the same authors as "computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer."
- **Reinforcement Learning:** An algorithm which learns to control and predict data. The algorithms are reward and goal orientated: "Reinforcement learning is learning what to do –how to map situations to actions– so as to maximize a numerical reward signal. The learner is not told which actions to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them" (Sutton & Barto, 2012). See also below for Deep Reinforcement Learning.

Machine Learning: The most prevalent of these narrow A.I. sub-domains, in an operational context, is Machine Learning. ML algorithm can be either supervised, unsupervised or semi-supervised. The majority of current ML implementations are supervised learning. In supervised learning, the idea is we (*humans*) teach the computer how to do something. In unsupervised learning the machine learns by itself (Samuel, 1959).

ML systems are being used to help make decisions both large and small in almost all aspects of our lives, whether they involve simple tasks like dispensing money from ATM's, recommendations for buying books or which movies to watch, email spam filtering, purchasing travel arrangements and insurance policies, to more objective matters like the prognosis of credit rating in loan approval decisions, and even life-altering decisions such as health diagnosis and court sentencing guidelines after a criminal conviction.

Systems utilizing ML information processing techniques are used for profiling individuals by law enforcement agencies, military drones, and other semi-autonomous surveillance applications. They capture information in our smart phones on our daily activities, from exercise and GPS data that tracks our location in real time, to emailing and social media interests and telephone calls. They are increasingly used in our cars and our homes. They are used to manage nuclear reactors and for managing demand across electricity grids, improving energy efficiency, and generally boosting productivity in the business environment.

Deep Learning: Deep learning is emerging as a primary machine learning approach for important, challenging problems such as image classification and speech recognition. Deep Learning methods have dramatically improved machine capabilities in speech recognition, approaching human-level performance on some object recognition benchmarks (He et al., 2016) and object detection (Ba, Mnih, & Kavukcuoglu, 2015). Which can also be very useful for self-driving cars and in many other domains where big data is available such as drug discovery and genomics (Nguyen et al., 2016).

Advances in Deep Learning will have broad implications for consumer and business products that can be significantly augmented by speech recognition. “Deep learning is becoming a mainstream technology for speech recognition at industrial scale” (Deng et al., 2013). This is particularly prevalent in telemarketing, tech help support desks (Vinyals & Le, 2015), and mobile personal assistants such as Apple’s Siri, Microsoft’s Cortana, Google Now, and Amazon Echo. Deep Learning is also being used for negotiations with other chatbots or people (Lewis et al., 2017).

Reinforcement Learning: Reinforcement Learning has gradually become one of the most active research areas in Machine Learning, Artificial Intelligence, and neural network research (Sutton & Barto, 2012). An RL agent interacts with its environment and, upon observing the consequences of its actions, can learn to alter its own behaviour in response to rewards received (Arulkumaran et al., 2017).

Within health, RL is being used for classifying gene-expression patterns from leukaemia patients into subtypes by clinical outcome (Ghahramani, 2015). These models have also contributed to massive savings at multiple Google Data Centers by helping to produce a 40% reduction in energy used for cooling and 15% reduction in overall energy overhead (Evans & Gao, 2016). Other typical examples of uses might include detecting pedestrians in images taken from an autonomous vehicle. As shown in (Shalev-Shwartz, Shammah, & Shashua, 2016), RL is proving to be especially effective in the development of self-driving cars which requires many capabilities such as sensing, vision, mapping, knowledge of driving policies, and regulations.

In robotics, RL is making progress in other seemingly simple tasks such as screwing a cap onto a bottle (Levine et al., 2016) or door opening (Chebotar, 2017).

A well-known successful example of RL is from the Google owned company DeepMind, specifically their AlphaGo, which defeated the human world champion in the game of Go. AlphaGo was comprised of neural networks that were trained using supervised and reinforcement learning in combination with a traditional heuristic search algorithm (Silver et al., 2016).

Deep Reinforcement Learning: One of the driving forces behind Deep Reinforcement Learning is the vision of creating systems that are capable of learning how to adapt in the real world. Further, researchers consider that “DRL will be an important component in constructing general AI systems” (Arulkumaran et al., 2017). As was shown through a single DRL architecture “in a range of different environments with only very minimal prior knowledge” (Mnih et al., 2015).

To date, DRL has been most prevalent in games (Mnih et al., 2013); however, recent developments have shown that DRL algorithms have by “far the most complex behaviors yet learned” in a machine algorithm (Christiano et al., 2017).

b) What factors have contributed to this? Historically, developments in A.I. were driven by government investment in research and development within academia and other research institutes. Whilst governments around the world still make large investments into A.I. research, recent major advances have largely been driven by significant investments by leading technology companies relying on techniques that were previously developed through government and other institutions investment.

Furthermore, computing power has increased dramatically. Meanwhile, the growth of the Internet and social media in the last 10 years has provided opportunities to collect, store, and share large amounts of data. Many leading technology companies are amassing huge amounts of ‘Big Data,’ supported in part by cloud computing resources. These companies have invested heavily in A.I. technologies and further seek to develop A.I. techniques to ensure a competitive advantage.

Another major factor is open access of scientific inventions and research in general –sites such as arXiv, provide immediate online publication of research papers, conference proceedings, etc. Additionally, open source frameworks and libraries for the development of ML algorithms have put opportunities for development into the hands of millions, thereby profiting from the advantages of cloud computing and parallel processing on GPUs. Examples include TensorFlow, Theano, CNTK, MXNet, and Keras. They implement model architectures and algorithms for methods, especially deep learning that can be run by calling functions without the need to implement them from scratch nor locally.

c) How is it likely to develop over the next 5, 10 and 20 years. There are several recent surveys of experts opinions on when A.I. will be available and their impact on the workplace. Many uncertainties exist concerning future developments of machine intelligence, one should therefore not consider the ‘expert view’ to be predictive of likely ten and twenty year scenarios.

d) What factors, technical or societal, will accelerate or hinder this development? There are some obvious factors such as a slow-down in

investment which would impact research and development and education, creating another 'A.I. winter' and skills gap. Other factors such as global instability and government policy, may all hinder the development of A.I.

Although the particular narrow A.I. models we outlined above already demonstrate aspects of intelligent abilities in narrow and limited domains, at this point they do not represent a unified model of intelligence and there is much work to be done before true A.I. is 'amongst us.'

Further, technically there are still many factors that make narrow A.I. unstable. Additionally there are technological challenges to overcome such as the curse of dimensionality –Richard Bellman (1957) asserted that high dimensionality of data is a fundamental hurdle in many science and engineering applications. He coined this phenomenon the curse of dimensionality, although recent developments in DRL have made some progress in addressing the curse of dimensionality (Bengio, Courville, & Vincent, 2013; Kulkarni et al., 2016). There are also many safety challenges to overcome such as security, data privacy (see for example (DeepMind, 2017)) and other technological problems still requiring breakthroughs.

Other advances will accelerate A. I. such as Facebook CommaAI (Baroni et al., 2017) and their A.I. roadmap (Mikolov, Joulin, & Baroni, 2015). Together with closer cooperation with Neuroscience and A.I. developers (Hassabis et al., 2017). We also believe the following papers will contribute to the acceleration of narrow A.I. solutions for mainstream uses beyond games and social media analytics: (Kalchbrenner, Danihelka, & Graves, 2015; Lake et al., 2016; Mnih et al., 2015).

2. We recommend the committee consider the findings in the paper by leading A.I. researchers at Microsoft, Ethan Fast and Eric Horvitz, *Long-Term Trends in the Public Perception of Artificial Intelligence* (Fast & Horvitz, 2017).

3. It is our belief that the goal of A.I must be to support humanity. At the present time it is difficult to predict the short term extent with which A.I. will impact on social and economic institutions but in the long term it could have a major negative consequence the social and economic effects of which could be severe for millions of people. In this case, according to a report to the US President of the United States (Furman et al., 2016), "Aggressive policy action will be needed to help (those) who are disadvantaged by these changes and to ensure that the enormous benefits of AI and automation are developed by and available to all." Other commentators such as Andrew Haldane (2015), Chief Economist at the Bank of England, believe it is clear that the introduction of AI machines and more advanced robotics could see a technological change and thus social and economic changes far larger than at any time in human history with massive unemployment of unprecedented scales.

Conversely, machines have been substituting human labor for centuries; yet, historically, technological changes have been associated with productivity growth and expanding rather than contracting total employment and with raising earnings. Research showed that factories that have implemented industrial robots also added over 1.25 million new jobs from 2009 to 2015 (Lewis, 2015).

The challenge for policymakers will be to update, strengthen, and adapt policies to respond to the social and economic effects of A.I. We have created an agenda with key research goals to ensure the development and the outcomes of A.I. and Artificial General Intelligence (AGI) are aligned with the social and economic advancement of all humanity, and how best to close those social and economic gaps through beneficial AI and AGI development.

4. Overall we believe that whilst some large corporations and their shareholders will benefit from the gains of A.I. the potential for artificial intelligence to enhance people's quality of life in areas including education, transportation, and healthcare is vast. However, we are willing to offer our expertise to the committee so that government, policy makers, and researchers collaborate to develop and champion methodology "for wealth creation in which everyone should be entitled to a portion of the world's A.I. produced treasures" (Stone et al., 2016).

5. Our research shows that theories of intelligence and the goal of A.I. have been the source of much confusion both within the field and among the general public. To help rectify this we are conducting a research survey: *Defining (machine) Intelligence* (Lewis & Monett, 2017).

The research survey on definitions of machine and human intelligence is still accepting responses and has an ongoing invitation procedure. However, we are incredibly surprised by the volume of responses together with the high level of comments, opinions, and recommendations concerning the definitions of machine and human intelligence that experts around the world have shared. As of September 6, 2017 we have collected more than 400 responses.

A.I. has a perception problem in the mainstream media even though many researchers indicate that supporting humanity must be the goal of AI. By clarifying the known definitions of intelligence and research goals of Machine Intelligence this should help us and other A.I. practitioners spread a stronger, more coherent message, to the mainstream media, policymakers, and the general public to help dispel myths about A.I.

6. We recommend the committee consider the findings projected through to 2030 in the report, *The One Hundred Year Study on Artificial Intelligence* (Stone et al., 2016), especially the sections on transportation, healthcare, education, low-resource communities, and public safety and security.

8. Human intellect is the source of many of its own problems. Errors in thinking and biases, which have grown powerful over time, are also showing up in the intelligent machines we program and may become even more prevalent in machines programmed with Artificial Intelligence.

Machines can no more do ethics than they can have psychological breakdowns. They can help to change circumstances, but they cannot reflect on their value or morality. It is the human element and bias that must be considered above all else.

9. For an 'unbiased' view see paper by Adrian Weller (2017) where he states "a brief survey, suggesting challenges and related concerns. We highlight and review settings where transparency may cause harm, discussing connections across privacy, multi-agent game theory, economics, fairness and trust."

The role of the Government

10. Key questions which governments and policy makers should be addressing are:

- How do we mitigate the uncertainty and likelihood of massive unemployment?
- What impact have A.I. systems and robots had in industrial factories? Have companies that employed robots, increased or decreased human employment?
- What new skills have been required as robots enter the workplace?
- Which new laws or modifications to laws will need to be implemented to mitigate risk and monitoring of A.I. and A.G.I.?
- Monitor and provide reporting on emerging technology policy, with a focus on artificial intelligence and automation.
- Provide research input into FLI's Asilomar long-term issues (Asilomar AI Principles, 2017) with particular focus on: "23) Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization."

References

- Arulkumaran, K. et al. (2017). A Brief Survey of Deep Reinforcement Learning. *CoRR*, abs/1708.05866, <https://arxiv.org/abs/1708.05866>.
- Asilomar AI Principles (2017). Future of Life Institute, <https://futureoflife.org/ai-principle>.
- Ba, J. L., Mnih, V., and Kavukcuoglu, K. (2015). Multiple Object Recognition with Visual Attention. *CoRR*, abs/1412.7755, <https://arxiv.org/abs/1412.7755>.
- Baroni, M. et al. (2017). CommAI: Evaluating the first steps towards a useful general AI. *CoRR*, abs/1701.08954, <https://arxiv.org/abs/1701.08954>.
- Bellman, R. (1957). *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press.

- Bengio, Y., Courville, A., and Vincent, V. (2013). Representation Learning: A Review and New Perspectives. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828.
- Calo, R. (2017). Artificial Intelligence Policy: A Roadmap, <https://ssrn.com/abstract=301535>.
- Chebotar, Y. et al. (2017). Path integral guided policy search. *CoRR*, abs/1610.00529, <https://arxiv.org/abs/1610.00529>.
- Christiano, P. F. et al. (2017). Deep Reinforcement Learning from Human Preferences. *CoRR*, abs/1706.03741, <https://arxiv.org/abs/1706.03741>.
- DeepMind (July 2017). What we've learned so far, <https://deepmind.com/applied/deepmind-health/transparency-independent-reviewers/what-weve-learned-so-far/>.
- Deng, L. et al. (2013). Recent advances in deep learning for speech research at Microsoft. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 8604–8608, IEEE.
- Evans, R. and Gao, J. (2016). DeepMind AI Reduces Google Data Centre Cooling Bill by 40%. DeepMind, <https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40>.
- Fast, E. and Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI-17*, San Francisco, CA, USA, February 4-9, 2017. AAAI Press, pp. 963–969.
- Furman, J. et al. (2016). Artificial Intelligence, Automation, and the Economy. Executive Office of the President, Washington, D.C. 20502, <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>.
- Ghahramani, Z. (May 2015). Probabilistic machine learning and artificial intelligence. *Nature*, 521:452–459. DOI: 10.1038/nature14541.
- Haldane, A. (2015). Labour's Share – speech given at the Trades Union Congress, London. Bank of England, <http://www.bankofengland.co.uk/publications/Pages/speeches/2015/864.aspx>.
- Hassabis, D. et al. (July 2017). Neuroscience-Inspired Artificial Intelligence. *Neuron*, 95(2):245–258.
- He, K. et al. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*. Las Vegas, NV, USA, pp. 770–778, IEEE.
- Hutter, M. (2005). *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Berlin: Springer.
- Kalchbrenner, N., Danihelka, I., and Graves, A. (2015). Grid Long Short-Term Memory. *CoRR*, abs/1507.01526, <https://arxiv.org/pdf/1507.01526.pdf>.
- Kulkarni, T. D. et al. (2016). Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation. *CoRR*, abs/1604.06057, <https://arxiv.org/abs/1604.06057>.
- Lake, B. M. et al. (2016). Building Machines That Learn and Think Like People. *Behav Brain Sci.*, 4:1–101.

- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature*, 521:436–444.
- Levine, S. et al. (January 2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(1):1334–1373.
- Lewis, C. W. P. (2015) Study – Robots are not taking jobs. *Robotenomics*, <https://robotenomics.com/2015/09/16/study-robots-are-not-taking-jobs>.
- Lewis, C. W. P. and Monett, D. (2017). Research Survey: Defining (machine) Intelligence. Ongoing survey, <https://goo.gl/hMjaE1>.
- Lewis, M. et al. (2017). Deal or No Deal? End-to-End Learning for Negotiation Dialogues. *CoRR*, abs/1706.05125, <https://arxiv.org/abs/1706.05125>.
- Mikolov, T., Joulin, J., and Baroni, M. (2015). A Roadmap towards Machine Intelligence. *CoRR*, abs/1511.08130, <https://arxiv.org/abs/1511.08130>.
- Mnih, V. et al. (2013). Playing Atari with Deep Reinforcement Learning. *CoRR*, abs/1312.5602, <https://arxiv.org/abs/1312.5602>.
- Mnih, V. et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518:529–533.
- Nguyen, D.-T. et al. (2016). Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Research*, 45(D1):D995–D1002.
- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3):535–554.
- Shalev-Shwartz, S., Shammah, S., and Shashua, A. (2016). Safe, Multi-Agent, Reinforcement Learning for Autonomous Driving. *CoRR*, abs/1708.05866, <https://arxiv.org/abs/1708.05866>.
- Silver, D. et al. (January 2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 28; 529(7587):484–489.
- Stone, P. et al. (September 2016). Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA, <http://ai100.stanford.edu/2016-report>.
- Sutton, R. S. and Barto, A. G. (2012). *Reinforcement Learning: An Introduction*. Second edition. London, UK: The MIT Press.
- Vinyals, O. and Le, Q. V. (2015). A Neural Conversational Model. *CoRR*, abs/1506.05869, <https://arxiv.org/abs/1506.05869>.
- Weller, A. (2017). Challenges for Transparency. *CoRR*, abs/1708.01870, <https://arxiv.org/abs/1708.01870>.
- Wiener, N. (1960). Some Moral and Technical Consequences of Automation. *Science*, 131(3410):1355–1358.

6 September 2017

The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB) – Written evidence (AIC0086)

Introduction

1. This submission, dated **Tuesday 5 September 2017**, is from Andrew Owen Martin, Secretary of The AISB [<http://aisb.org.uk/>] on behalf of the members. Contributions were received from;

- 8 the Chair, Berndt Müller
- 9 the Treasurer, RobWortham;
- 10 members, Ruth Aylett,
- 11 Annemarie Naylor,
- 12 and Martin Wurzinger.

2. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB) is the largest Artificial Intelligence Society in the United Kingdom. Founded in 1964, it is the oldest society for the study of AI in the world, and has an international membership drawn from both academia and industry.

3. The evidence herein pertains to the following

- Paragraphs 4-7 A description of the state of AI which is as accurate today as when it was authored in 1973.
- Paragraphs 8-12 The single issue which has halted the progress of all AI projects.
- Paragraphs 13-17 The bureaucratic nature of AI systems.
- Paragraphs 18-20 A view of the effect of automation on society.
- Paragraphs 21-26 A proposal of a policy for the design of AI systems.
- Paragraphs 27-29 A solution for the problem of data-based monopolies.
- Paragraphs 30-37 Summarising the evidence of previous sections.

On the definition of Artificial Intelligence

4. In this document, the term Artificial Intelligence will be used as used by Prof. Sir James Lighthill in his report, commissioned by the Science Research Council (SRC), to give an unbiased view of the state of AI research in 1973.
 5. Lighthill was able to place all work on AI into one of three categories. Category **A**, for **A**dvanced **A**utomation and **A**pplications. Category **C**, for **C**omputer-based **C**entral Nervous System research. Category **B**, for **B**uilding Robots, and for any work acting as a **B**ridge between categories **A** and **C**.
 6. In the report, work in category **A** was identified as making rapid progress, and achieving commercial success. Work in category **C** was also making progress, and producing all the benefits associated with scientific discoveries. Work in category **B** was identified as being characterised by projects which made grand claims about being able to expand success in a small task into the solving of a whole scientific domain, claims which were never realised.
 7. It is the view of this document that this definition of the field of AI is as relevant and accurate in 2017 as it was in 1973. The implications thereof are that work identified as category **A** or **C** should be treated as any other progress in science or engineering, which historically brings improved productivity, and social unrest. Work which aims to duplicate human intelligence by bridging categories **A** and **C**, either by reproducing a central nervous system on computers, such as the Human Brain Project, or by an exploration of algorithms which produce generally intelligent behaviour, should be approached with scepticism about the grandiosity of the claim, and with a sharp focus on the limited scope of previous success.
- Q2: Is the level of excitement warranted? On the inability of AI systems to generalise their knowledge
8. Historically, AI projects have demonstrated impressive results in their early stages, and have always failed to extend that success into a wider area.
 9. Without going into too much detail, the history of AI is littered with examples of a research time selecting a problem, such as being able to respond to commands typed in English, and achieving success in a limited domain, such as when commands are limited to using only words from a specific list, then claiming all of English will be understood by the system in the near future, and never being heard of again. Famous examples are Newell and Simon's General Problem Solver, Winograd's SHURDLU, Lenat's Cyc, Brooks's Cog.
 10. The principle is that it is relatively easy to design a set of rules for behaviour in an area, for example a game, a language, or a job, if there is a very clearly defined context, such as strong opening moves in the game of Go, using a fixed word list in a language, and simply scanning products in at checkout, but that no set of rules will ever cover all situations that could emerge, such as your opponent having a novel strategy for Go, creative use of a language, or suspecting a theft at a checkout. It is a reasonable observation that the whole of

Go is quite a well defined context, being fully described in fewer than one hundred rules, and indeed AI has recently made impressive progress in this area, however nothing in human life is as regular and clearly defined as Chess.

11. In 2015, the AI company DeepMind demonstrated a system which could learn to play Atari video games, this was hailed as the first step towards general learning systems, even though their system uses the latest techniques, history has told us this is not so.

12. We are concerned that there is no distinction made between the hugely speculative opinions on the impact of AI such as those which predict an apocalypse or utopia, and those opinions which are historically justified such as the impact of automation. **We therefore recommend** that any project requesting government funding must declare a well defined scope of the project or otherwise justify how it will avoid the problem of all AI systems, that it cannot generalise its knowledge to situations for which it was not explicitly designed.

Q1: On the current state of AI systems, their fundamentally bureaucratic nature

13. All AI systems are limited in the same way, they can only follow their rules. This has two big implications, (i) an AI system must have its task defined in explicit rules, and (ii) an AI system will never be able to judge when its rules do not apply.

14. Neural networks are AI systems which generate appropriate behaviour from the connectivity of a large number of simple processing nodes. While these systems produce some of the most impressive behaviour, most notably in the cases of IBM's Jeopardy! playing Watson and DeepMind's Go playing AlphaGo, they still merely follow the rules of the algorithm which simulates the network of nodes. This adds the problem that these systems provide no explicit meaningful rules which can be traced back to help understand what lead to their decisions.

15. To illustrate these problems, imagine a computer to be a very quick, but utterly ruthless bureaucrat. Many people have had the experience of having to answer a question on a form that does not apply to them, then having to judge if answering incorrectly is worse than not answering at all. Even in the world where intelligent people are processing the forms this can cause problems, but mostly a person can understand when answers are not relevant, but these problems are exacerbated if the forms are processed by a fanatic bureaucrat, or machine.

16. Similarly, when a bureaucratic system is designed to evaluate someone's behaviour it always risks defining their behaviour. Take for example when a customer services employee is paid by the number of phone calls they answer, they may be encouraged to immediately answer and hang up on any phone call they receive as they are not judged on any other aspect of their performance. This is not such a problem when a person is able to review the whole situation, but if they can only see a report, and this is the only information an AI system will receive, they will highly reward this behaviour.

17. Running a bureaucracy entirely on a computer can speed it up, but it will not solve the problems of people exploiting the loopholes, or the rules failing to capture the true complexity of the system. **We therefore recommend** that either (i) a human supervisor is employed to accept or reject all decisions made by an AI system, or (ii) the person who decided to deploy the AI system, be named and held legally accountable for all decisions made.

Q3: Preparing for the impact of AI on society.

18. Progress in category **A** AI should be seen as the introduction of new tools to society. Tools share a common impact trajectory, whether they be the hand plough, cotton spinning and weaving machines, or self-driving cars. The human skills required before they came along decrease in value. At the same time, new skills are enabled, creating opportunities for those in a position to exploit them. In some cases the skill will not be missed, mostly in jobs constituted by rigorous rule following, such as human calculators or elevator operators. In other cases the devalued skills may benefit from being protected to prevent cultural loss.

19. When a job is automated, there are always some unidentified parts of the job lost, in the case of the elevator operator they may have provided a valuable service of recognising vulnerable or suspicious elevator riders, human computers may have provided a service of indicating when they were being asked to make the wrong calculations. People displaced from their jobs by automation are the perfect people to recognise the limitations of their replacement.

20. **We therefore recommend** to (i) recognise which skills are being displaced and to conserve those of particular cultural value, and (ii) provide support for the people who have had their skills devalued and those intending to exploit new opportunities who might otherwise be overshadowed by existing interests, and (iii) consider experience in a job as valuable qualification for supervisory roles of automated versions of that job.

Q10: On the role of the Government

21. All robots are designed by humans, and it is the designer's responsibility to make robots that are not mysterious or difficult to understand, and to make them transparent. Robots are tools, not mechanical people, even if they can perceive and interact with the world like humans. We need our robots to constantly remind us of this fact by explaining themselves to us as we encounter them, and as they interact with us. Similarly, robots should also be required to identify themselves to other robots to avoid "confusion" amongst machines.

22. The UK currently leads the world in this area, as we already have a set of guidelines, the 'EPSRC Principles of Robotics'. These are guidelines for robot designers, manufacturers and operators. But they are only guidelines, and as yet they are not enforced. Other bodies like the IEEE are working on a global initiative to set standards for AI and autonomous systems.

23. The EPSRC Principles of Robotics may be summarised as follows:

1. Robots should not be designed as weapons, except for national security reasons.
2. Robots should be designed and operated to comply with existing law, including privacy.
3. Robots are products: as with other products, they should be designed to be safe and secure.
4. Robots are manufactured artefacts: the illusion of emotions and intent should not be used to exploit vulnerable users.
5. It should be possible to find out who is responsible for any robot.

24. Building transparent AI is one way to help designers and users of complex AI systems to better understand, and thus calibrate their trust in them, and interactions with them.

25. Rob Wortham has given a TEDx talk on this subject <https://www.youtube.com/watch?v=st65KZkC3qM>.

26. **We therefore recommend** the Government takes its lead from these standards initiatives.

Q7: On data-based monopolies

27. Future Care Capital recently published a research report, Intelligent Sharing: unleashing the potential of health and care data in the UK. The recommendations were as follows.

- Increasing investment and support for data controllers to unleash health and care data in a standard and anonymised form - where there is value in secondary analysis by third parties;
- Establishing a new National Health and Care Data Donor Bank to coordinate data from the public and help improve the alignment of research to clinical need; and
- Cross-sector coordination of efforts designed to stimulate a culture of data philanthropy by individuals, communities of interest and corporates alike.

28. Some have called for greater use of existing powers and sanctions, whilst others have called for new anti-trust legislation, to address data-based monopolies in the interests of stimulating innovation and enterprise. By contrast, and in recognition of the very significant personal data holdings such monopolies represent, we have suggested that Government should explore the development of a gift aid style scheme for health and care data and encourage individuals to make health and care data donations to further promote the reuse of corporate data to transform related outcomes - or what has been termed the use of data for good.

29. A full copy of the report is available here:

<https://futurecarecapital.org.uk/policy/intelligent-sharing-unleashing-the-potential-of-health-and-care-data-in-the-uk-to-transform-outcomes/>

Conclusion

30. This submission represents the views of Andrew Owen Martin, the secretary of the AISB and supported by contributions from the chair, the organising committee, and members.

31. Lighthill's report on these issues was as astute then as it is now. Computer power and techniques have changed, but the principles have not.

32. AI is not human intelligence, human intelligence is not AI. AI is an evocative term, but the history of AI systems failing to work outside of tightly defined contexts shows that they are fundamentally distinct from human, or even animal, intelligence.

33. AI systems are akin to employing a very fast, but fanatical bureaucrat to run a company or department. There are well-known problems of bureaucracy which are magnified in the case of AI. A human must be legally accountable for the decisions of an AI system.

34. We should expect the skills of individuals to be devalued, be prepared to support them, possibly by employing them as supervisors of machines performing their skills, and take action to preserve certain human skills from being lost altogether.

35. The existing guidelines for building robots proposed by the EPSRC should be adopted and continuously developed further.

36. Future Care Capital has a proposal to of a Gift Aid style scheme to encourage individuals to make donations of their health and care data. We support this proposal.

37. In conclusion AI will affect society in the same way that technology has done throughout history, no one can tell now how much each industry will be affected in the future, but we can be prepared to act upon two historical principles; (i) people who lose their jobs due to technological progress will need support, and (ii) there is no humanity in "only following orders" which is all machines can do.

5 September 2017

The AI Initiative, The Future Society at Harvard Kennedy School – Written evidence (AIC0209)

Marie-Therese Png, Esam Goodarzy, Imaan Binyusuf, AI Initiative at Harvard Kennedy School

The AI Initiative is an initiative of The Future Society at Harvard Kennedy School dedicated to the rise of Artificial Intelligence. Created in 2015, it gathers students, researchers, alumni, faculty and experts from Harvard, MIT and beyond, interested in understanding the consequences of the rise of Artificial Intelligence. Its mission is to help shape the global AI policy framework.

1. Defining Artificial Intelligence

According to John McCarthy, who coined the term 'Artificial Intelligence' in 1955, AI is "the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."²⁵

A key distinction within AI lies between Artificial Narrow Intelligence (also called "weak" AI), and Artificial General Intelligence (also called "strong" AI). Artificial General Intelligence refers to an autonomous machine's ability to perform any intellectual tasks that a human can perform. Artificial Narrow Intelligence refers to an autonomous machine which operates strictly within the confine of the scenarios for which they are programmed.²⁶

Nonetheless, definitions of AI underestimate the complexity of human intelligence in terms of the ability to process and apply information across a wide range of natural and abstract domains. Thus, to date only narrow AI has been developed, specialised in small breadths of intelligent behaviour - Artificial General Intelligence remains elusive. Further, the common assumption that human intelligence can be replicated by precisely modelling computational structure of a biological brain may be misled - intelligence is also a socio-cultural phenomenon. Hence the expected arrival date of human level machine intelligence is receding at a rate of one year per year.

²⁵ <http://www-formal.stanford.edu/jmc/whatisai/node1.html>

²⁶ Making the AI revolution work for everyone, March 2017 The Future Society, AI Initiative Nicolas MIAILHE Cyrus HODES

2. How can the general public best be prepared for more widespread use of artificial intelligence?

Two key issues we would like to focus on is the digital divide, and algorithmic bias.

The AI Digital divide is a term popularised with the advent of the internet in the 1990s. As the global diffusion of the internet emerged, the digital divide referred to the inequality of access between the 'haves' and 'have-nots'.²⁷ This dichotomous perspective informed early UK digital inclusion policies, with the normative assumption that the digital divide could be overcome by simply improving digital infrastructure and broadband connections. This digital divide will be carried through, if not exacerbated by the exponential changes AI can entail. These inequalities can be measured using five dimensions developed by DiMaggio and Hargittai²⁸, to explain the differential use of new technologies. This framework outlined access to equipment, autonomy of use, skill to use, purpose of use, and social support as essential factors where inequality arises.²⁹ As such, these dimensions should be considered in constructing interventions through which to prepare the public for more widespread use of artificial intelligence.

The second issue is algorithmic bias. Machine learning tools are often positioned as fair and objective. Today, they enable us to translate languages, recognise faces, emotions, objects, and speech. However, algorithmic bias, like human bias, can result in exclusionary experiences and discriminatory practices.³⁰ Algorithms are often based on both historical precedent, frequently reflecting the majority. AI is a technology which learns fast using examples. As technology becomes increasingly complex and ubiquitous, this bias becomes further obscured while continuing to scale. The algorithms are neither accessible nor transparent, and as a result, are ripe with unintended consequences. This can create a destructive feedback loop that support a narrow interpretation of identity and reality that are often functions of proxies rather than fact.

It has become increasingly apparent in research, as well as political discourse, that the algorithms upon which AI is based on may contain certain biases. It has been explained that algorithms are trained on data sets which are reflective of the inequities present in society, whether they be socioeconomic, racial, gender, ability, religion, sexual orientation, etc. For example, certain risk assessment

²⁷ Bostock and Steptoe, 2012

²⁸ DiMaggio and Hargittai (2001)

²⁹ Inequality.co.uk: Exploring UK Digital Inequality for young people seeking online sexual health information Imaan Binyusuf, 2017

³⁰ See <https://www.nytimes.com/2015/08/11/upshot/algorithms-and-bias-q-and-a-with-cynthia-dwork.html?mcubz=3> and <http://ajlunited.org>

software used in criminal sentencing has been reported to be being racially biased.³¹ As such, discussions of bias and discrimination have taken on a renewed sense of urgency, notably regarding the use of AI to analyse data to inform government decisions.

3. What are the ethical implications of the development and use of artificial intelligence?

We are transitioning towards a world in which our decisions of how to gather, organise, and act on information are increasingly being governed by systems that are opaque to most, if not all of us. Ensuring that these systems work towards our collective well-being requires us to be exceptionally lucid in our representations of the values we hold and the ways we want those values to be optimised for and/or preserved. Our ability to engage with philosophy and converge on ethical principles is starting to take on a dimension of significance that has not existed before.³² For example, western philosophy has referred to a thought experiment called the “Trolley Problem” for decades, which describes a situation where a subject would be standing at a junction of rails deciding whether to flip a switch causing a trolley to take the life of an innocent bystander in order to avoid killing five people that the trolley is heading towards. This thought experiment is analogous to actual scenarios that automated vehicles will inevitably face as they start to become integrated onto our roads. Complex high-stakes scenarios will arise, and it remains to be seen whose value systems will be behind the engine of the artificial intelligence (AI)’s algorithm.

Converging on our values within this space and representing those value systems within the AI’s architecture are both distinct and important challenges to overcome. However, even if we had agreed upon those value systems and represented them within the AI’s algorithms perfectly, the potential ethical issues that arise can be insidious. Most notably, researchers in the field of AI value alignment have limited understanding of how to build in values to an AI’s architecture without the AI compromising other values we hold but couldn’t possibly foresee being undermined.

Currently, there is at least a 10:1 ratio of researchers who are working in AI development to researchers who are focused specifically on AI safety. For this reason, we need more resources, both in the form of funding and people, dedicated to the field of AI safety research, as is being worked on by organisations such as the Future of Humanity Institute and the Machine Intelligence Research Institute.

³¹ See <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

³² See <https://intelligence.org/2016/12/28/ai-alignment-why-its-hard-and-where-to-start/>

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

The Future Society’s AI Initiative endeavours to engage multi-stakeholders to shape the global policy framework needed to harness the opportunity and address the challenges of the development and control of Artificial Intelligence. We will be hosting the Global Civic Debate on the governance of AI over the course of the next few months through a collective intelligence platform³³ with the aim to provide holistic policy solutions in a rapidly transforming field.

The Future Society, the IEEE, the JSAI and bluenove are joining hands to launch and lead, in the next 6 months (September 2017 - February 2018), a global civic debate on “Governing the rise of AI”. The consultation is open to everyone. It will rely on the award-winning collective intelligence platform “Assembl” and associated methodology developed by bluenove, a pioneer in the field of civic tech. Assembl has already been deployed successfully in several public debates in the past on a host of complex topics including Smart Cities, the rise of inclusive cities, the future of education, the soft power of expatriates and the OECD well-being index.

Led over a period of 6 months by the “AI Initiative”, the civic consultation will seek to proactively involve citizens, practitioners, world experts, and researchers working on AI, robotics, cyber, public policy, international relations and economics. This civic consultation will proactively engage citizens, practitioners, world experts, and researchers working on AI, robotics, cyber, public policy, international relations and economics. The aim of the civic debate is to trigger a broad and inclusive dialogue which informs our understanding of the dynamics and consequences of the rise of AI and how to govern the current technological revolution. The collective intelligence platform will lead to a synthesis of the main ideas and proposals, which will in turn be used to inform solutions and actionable policy tools for the international governance of AI.

The debate will take place in 4 phases:

DISCOVERY	IDEATION	EXPLORATION	CONVERGENCE
September 7th – October	October– November	November– December	January– February 2018
Where quick thoughts are expressed on 4 main topics: <ul style="list-style-type: none"> • The AI Revolution • AI for the common good • AI Impact on the workforce • Possible futures by 2045 	Subjects will be then discussed in detail through an argumentative forum.	Will tackle the most promising ideas emerging from the debate.	Where the best ideas will be selected for implementation.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so- called ‘black boxing’) acceptable? When should it not be permissible?

³³ <https://assembl-civic.bluenove.com/ai-consultation/home>

Redefining accountability when an AI decision process is largely known and perceived as neutral creates a significant obstacle to guiding AI for positive public service. Applying machine learning to public policy for increased quality of life, is an overt interaction with the social. Large scale impacts of AI will inevitably have ethical ramifications, and thus AI technology cannot be framed as neutral or apolitical. If AI is perceived as objective, its decisions will define epistemological systems and “truth”. Technologies have the power to shape and absorb the values of dominant individuals, companies, or institutions. AI as decision making technologies, when perceived as neutral, falls as a knife onto a grey area and turns it black and white, not only from thereon, but also retrospectively.

This is particularly concerning with regards to the use of AI and Social Media in political Campaigning. Persuasive technologies such as Facebook, shapes the thoughts and beliefs of billions of people. It is only recently that it has been unveiled as politically salient, rather than a public utility, with the assertion that Cambridge Analytica used Facebook data to facilitate Trump’s rise to power. Thus, AI is identified as a precipitator of political, but also cultural, and socio-economic shifts. Concerns around bias, transparency and autonomy, in AI and wider society, are increasingly entering the public discourse. My future career will concern the ethical analysis and application of machine learning and neural networks in public policy, with the goal of increasing the quality of life of wider populations.

Though there are many potential benefits, there is an increasing conscience of biases in data used to train an algorithms, reflecting a history of discrimination in recidivism risk, insurance, housing, hiring, medical and academic admission softwares are already biased towards specific populations(Kleinberg, 2016; Dwork, 2013; Kleinman, 2017). This is largely due to the “black box of machine learning”, where machine learning classifies data with coherent output, but even data scientists behind the code do not know the mechanism or reasoning behind the outcome (Kleinberg, 2016). Lines of code interact in ways that humans cannot make any judgement of bias. Therefore, transparency is critical- when a model gives a recommendation, we must understand upon what assumptions the decision is made, which are largely shaped by institutionalised social inequities.³⁴

11 September 2017

³⁴ Principles for Algorithmic Transparency and Accountability by the Association for Computing Machinery
US Public Policy Council (USACM) (2017)

The Alan Turing Institute – Written evidence (AIC0139)

The Alan Turing Institute makes this submission as part of the inquiry lodged by the House of Lords Select Committee on Artificial Intelligence.

The Alan Turing Institute is the UK's national institute for data science. Five founding universities – Cambridge, Edinburgh, Oxford, UCL and Warwick – and the UK Engineering and Physical Sciences Research Council created The Alan Turing Institute in 2015. Our goals are: to undertake world-class research in data science, apply our research to real-world problems, driving economic impact and societal good, lead the training of a new generation of data scientists, and shape the public conversation around data. Our researchers come from a wide variety of intellectual and disciplinary backgrounds, ranging from computer science, statistics, and mathematics, to social sciences. To reflect this, answers to questions are attributed to individual authors, and opinions may not always align. We also encourage the Committee to consider this evidence alongside that submitted by our five founding universities.

The pace of technological change

Question 1: What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Theo Damoulas (Turing Fellow):

1. AI will continue to rapidly advance over the next decades in both an inward direction (towards Strong AI/Artificial General Intelligence) and an outward direction (impacting science, industry, society, governance). These directions are complementary and create fertile ground and a positive feedback loop for AI advancement.
2. AI is already leading a revolution in analysing, understanding and optimising operations and governance in our cities. Many research initiatives under the umbrella themes of "smart cities", "urban informatics", "computational social science" and "data science centres" have sprung up around the world (MIT, NYU, Warwick, Glasgow, Santa Fe, The Alan Turing Institute, IBM, Microsoft, etc.) over the last 5 years. Areas that were traditionally in the remit of the social sciences such as understanding social/group behaviour and policy evaluation are and will continue to be transformed by AI.
3. In the next decades we will move from evidence-based policy informing to more active and integrated human-machine policy making and control. As we continue to improve our sensing capabilities of social systems, our algorithms and statistical models, our computational capabilities and AI systems, we will be able to understand better these complex, dynamic,

non-stationary systems. This will create unique advantages and opportunities for the social systems and governments that will reach this exciting nexus point. The UK has a unique advantage in reaching that point first through national initiatives such as The Alan Turing Institute, and a rapidly developing research capability in universities.

Simon DeDeo (Visiting Researcher at The Alan Turing Institute):

4. Artificial intelligence is the use of computers to predict, make decisions, and take action in the absence of explicitly-specified rules from a human programmer. While computers have been used to aid in complex decision-making for decades, they've mostly done so by implementing human-coded decision rules. As such, these older machines fall victim to what Alan Turing referred to as "Lady Lovelace's Objection": they could do only what we could specify in advance. The fly-by-wire system in a commercial aeroplane of the 1980s is an example of a non-AI system: while it is enormously complex, its audited if-then structure is not fundamentally different from the feedback loop of a thermostat.
5. Artificial intelligence is a fundamental advance on these older systems. Increases in computer power, new algorithms, and vast amounts of data have brought a new kind of machine into being. These systems make decisions on the basis of rulesets that have not been specified in advance. Advances in machine learning amount to discovering particularly fertile ways to constrain the space of rules the machine has to search, or in finding new and faster methods for searching it. As heuristics are stacked on top of heuristics, the downside can be to make the rules more tangled and harder to interpret than before.³⁵
6. It is recent developments in computer power that make the benefits of AI apparent: the same algorithms run on 1950s, or even 1990s-era computers, would be intellectual curiosities. Paired with 21st Century technology, however, they have the potential to transform the material, social, and political landscape; as economists, political scientists, philosophers, and workers in the field itself often suggest, they have the potential to alter the basic rhythms of human life in a fashion last seen at the beginning of the Industrial Revolution.

Ricardo Silva (Turing Fellow):

7. We expect to see a bigger role of autonomous systems in experimenting with their environment for the benefit of users. We already observe this phenomenon in modest ways, such as in websites that optimise their users' experiences by suggesting and presenting different combinations of website configuration, learning by observing the users' reactions. We shall see such systems gaining an increasing role on assisting scientists,

³⁵ See "Wrong side of the tracks: Big Data and Protected Categories". Simon DeDeo. Chapter in Big Data is Not a Monolith (MIT Press). Edited by Cassidy R. Sugimoto, Hamid Ekbia, Michael Mattioli (Available at: <https://arxiv.org/abs/1412.4643>)

engineers and managers in new discoveries and ways of solving their short-term and long-term tasks. For that to succeed, machines must be able to better communicate their way of thinking when engaging in problem solving, and professionals should learn to play to the strengths of machine-made suggestions.

Maria Liakata (Turing Fellow):

8. Natural Language Processing (NLP) is a field closely linked to artificial intelligence that studies computational methods for automatically identifying structure in human language, in order to perform various tasks.³⁶ These tasks assume some knowledge of language and range from relatively simple ones such as automatically recognizing entities in a text, to much more complex, yet fundamental ones, such as inferring the syntactic or semantic relations in a sentence (parsing), or automatically translating a text from one human language to another (machine translation). Technologies like Google translate, with which we interact every day, are based on NLP technology. Recent advances in deep learning have made us much better at addressing these tasks, as long as we have access to large amounts of data.

Question 2: Is the current level of excitement which surrounds artificial intelligence warranted?

Adrian Weller (Turing Fellow):

9. Recent advances in perception from deep learning are justifiably causing excitement, much of which is due to increased computing power and larger data sets. These trends are likely to continue, with significant energy and resources currently being invested into developing these further. Even if further development of these is slow, there are likely to be great benefits of using this increased computing power and data sets to improve medical diagnosis and transport planning. However, there are misconceptions among the general public about the narrow limits of current AI systems. We are still a long way from a general learning system with human-like intelligence that can acquire knowledge from one domain and apply that flexibly in others. Certainly researchers are actively working on these challenges, and it is conceivable that we are not many years away – but we might easily be many decades away.

Maria Liakata (Turing Fellow):

10. Both AI and Natural Language Processing are moving at a fast pace, and in the next few years it is possible we will have trained computational models that are much better at making use of context and background knowledge, and diverse sources of linguistic and other information such as images, to allow better inference mechanisms and common sense

³⁶ See Jurafsky and Martin 2009

reasoning. This will allow us to automatically interpret faster and draw more sophisticated conclusions. However, researchers in this area are also increasingly concerned about introduction of biases in the data. Problems stemming from the latter were famously depicted by Microsoft's chat robot Tay, which was supposed to learn to speak in the language of a teenager, based on data posted to it by humans, but ended up being shut down as it was learning to swear and propagate dangerous views.

Impact on society

Question 3: How can the general public best be prepared for more widespread use of artificial intelligence?

Helena Quinn (Policy Officer):

11. Data science skills are likely to continue to grow in demand. In parallel, businesses will need to be data literate in order to take advantage of artificial intelligence. The Alan Turing Institute offers breakfast briefings and executive education to businesses to help fill this gap in data literacy amongst senior business figures. The general public would benefit from online training and much greater emphasis on numerical and computing skills from school level and onto university, and training students from non-mathematical and physical sciences effectively in quantitative methods.

Josh Cows (Data Ethics Researcher):

12. The Alan Turing Institute is delighted to be partnering with the Nuffield Foundation on the recently announced Convention on Data Ethics, set to launch in 2018. The Convention will serve as a focal point for representatives of all sectors of society to collaboratively tackle the core ethical challenges posed by the rise of artificial intelligence, while engaging the public with these debates and their implications.

Question 4: Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Maria Liakata (Turing Fellow):

13. AI can lead to a very useful set of tools that could help us advance as a society. We could be better researchers and scientists by having more advanced search and information extraction mechanisms; it could help us look after the environment through more efficient use of sensor information; it could allow better understanding of health problems and help provide cures by combining diverse sources of information and offer cost beneficial and regular monitoring of health conditions, providing extra evidence to doctors in their assessment; it could help us govern our countries more democratically by allowing access to multiple views and

better understanding of people's stance to various policies; it could help provide diverse means for education and training in various fields; it could help reduce the amount of bureaucracy and routine jobs, allowing us to be more creative with our time and focus on important endeavours. However, for all this to happen we need to have highly educated citizens who can comprehend the benefits of this technology and make the most of it rather than being passive consumers. I think there is a serious danger in focusing on cost benefits from AI and compromising on human training. It is a mistake to be replacing jobs without promoting education; we need human experts rather than relying on systems as the experts. To improve our experience as a society it is important to invest in education to have citizens that can make the most of the new technology in their everyday life and improve human-human interaction.

Public perception

Question 5: Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Ricardo Silva (Turing Fellow):

14. One way of demystifying the use of AI is to see it as a toolbox for scaling up our potential. It fits the historical trend of using machines to solve tasks that otherwise would not be feasible or too costly. The real danger is society as a whole failing to train people to take new roles, not that we will run out of human potential. Access to proper training and information on how to make use of such resources is what the public should look forward to, what they should be incentivized to do, and what should be democratized.

Nathanaël Fijalkow (Research Fellow):

15. The issue of trust from the general public induces the most important challenge for the years to come: to make artificial intelligence trustworthy. To achieve this goal, the government can help in two different ways. First, it can, mobilise researchers and practitioners around a number of aspects crucial to the development of usable technologies: privacy, security, reliability, transparency. Great results have been obtained through research in understanding the mechanisms behind artificial intelligence. We need an increased effort to make this technology usable. Second, it can set standards, helping and encouraging companies to use artificial intelligence in a safe and responsible manner.

Industry

Question 6: What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Ricardo Silva (Turing Fellow):

16. AI will move to change more specialized, knowledge-rich, jobs. As a concrete example, we should expect major advances in data-centric engineering, where intelligent systems will help in the design and monitoring of many complex systems. The scope of this enterprise will be massive, varying from the facilitated creation of energy-efficient technology to the design of smart urban infrastructure, sensitive to disaster management and prevention. Engineers should be trained early on how to maximise their potential by engaging with intelligent systems design tools.

Nicolas Guernion (Director of Partnerships):

17. AI, and in particular machine learning, also sometimes referred to as 'soft AI', has the potential to transform all economic sectors, from retail with the development of more powerful recommender systems which can better address customer needs, through to insurance where it can enable more accurate risk prediction (such as via analysis of satellite images or applications of voice analytics), health and well-being (early prediction of disease), finance (fraud detection applications) and law enforcement (faster text data analysis).

18. Whilst all sectors stand to benefit from AI and data science, there are vast inequalities in how fast economic sectors can reap the benefits from advances in this fast-moving field. Some of the largest sources of inequalities are access to trained people in the field and infrastructure and data readiness. Large corporations such as Google, Apple, Facebook and Amazon (GAFA), as well as companies leading the gig economy revolution (e.g. Uber, Airbnb, etc.) are able to recruit the best and often price out the competition. Finance and retail are also building up their capabilities at a fast pace, leaving huge gaps which cannot easily be filled in other sectors, such as traditional manufacturing. Lack of infrastructure and data readiness in some sectors also compound this issue e.g. farming, bulk manufacturing.

Question 7: How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Simon DeDeo (Visiting Researcher at The Alan Turing Institute):

19. Privileged access to massive data sets in private hands is a basic source of power for corporations like Google, Facebook, and Amazon, and there are current incentives for that data to be highly restricted. A visitor to the data centres of a company such as Google is immediately struck by the extreme levels of security in place, in some cases comparable to what one might expect at a nuclear power plant. Theft of information from these databases is, under the current legal system, economically catastrophic.

20. But these restrictions are completely artificial and may be neither just or fair. The data that comprises such a large part of Google's or Facebook's stock of capital was created by the users of that system. Companies get this data almost invariably free-of-charge, from ordinary citizens – at best in implicit exchange for services such as a free e-mail address or social media identity. Yet opting out of an e-mail address or, increasingly, a LinkedIn or Facebook account, is no longer possible if one wants to hold a job. As argued by a number of researchers in the field, such as the virtual reality pioneer Jaron Lanier, citizens may well have an expectation that they can reap the value of their own labour in more direct forms.
21. This becomes even more clearly apparent when, for example, a company like Uber considers using the records of its drivers to train an AI system that will replace them, or Google uses the publicly available translations from the European Union's parliament to train automatic translation systems that then replace the people that created the translations in the first place.
22. Data monopolies may not only be unfair and unjust, but also economically counterproductive: there is enormous value to the human race to be uncovered by disinterested investigation into proprietary data streams by scientists and the general public – not to mention competing corporations. And the need for companies to protect their data against theft is liable to become an increasing liability for economic stability. One solution to these linked questions of justice, innovation, and economic stability is to consider a data-sharing "patent" system, or term-limited monopoly comparable to the patent system for inventors. Such a "data patent" system would have to reward both the companies that assembled these databases, and, crucially, the individual users that participated in their construction. Although it could be based on the inventor's patent, it would have to go beyond it in novel and innovative ways, including the possibility of cash micropayments or annuities to citizens who provided the eyes, hands, and minds that trained a company's algorithms.

Maria Liakata (Turing Fellow):

23. The potential positive outcomes of AI could become a problem for society if access to AI technology and the data which feeds into AI systems were available to or controlled by very few. For example, while personal computers have been main-stream for the past 15-20 years and most people would be responsible for their own data we are moving to a time where data is stored somewhere on the cloud and computers are merely virtual machines. This would mean that we are giving away a lot of personal freedom to whomever controls these technologies. To prevent this from happening we need the government or organisations trusted by the public rather than large monopolies to be handling this technology. Even better, individuals should be able to control and make use of their own data. For example, this would be invaluable in personalized health

monitoring, where individuals would decide who has access to their data and the technology that automatically processes their data.

24. The government could help against technology monopolies by large corporations by providing better incentives to smaller companies, making work at Universities and the public sector more appealing, and better funding research and technological development in public organisations. Another possibility would be to promote crowd-funding and longevity for open source projects as well as clarity in terms of the potential gains for citizens.

Ethics

Question 8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

Ricardo Silva (Turing Fellow):

25. Some AI systems will be required to run experiments with human subjects, which may be of unclear ethical viability. For instance, when testing pedagogical approaches with children in an intelligent tutoring program, or choosing environmental controlling procedures that may affect users' mood, it may be tempting to make the procedure as least intrusive as possible. However, we must ensure that appropriate permission for such experiments is obtained from the individuals concerned, and that the request for permission is communicated in a clear and unambiguous way.

Adrian Weller (Turing Fellow):

26. As AI systems are increasingly used to help make important predictions or decisions affecting our everyday lives (e.g. criminal sentencing, hiring decisions), we must take care to ensure that appropriate levels of trust in these systems are justified. This requires that we consider issues of fairness, transparency, privacy, reliability, security and value alignment. For any particular application, various groupings of these issues will be important in different ways. Hence, while we support calls for overarching principles, it will be important also to examine how these might apply in specific contexts. As one example, the use of personal data to decide health insurance premia worries many people, whereas a similar use to decide what to charge for car drivers' insurance seems reasonable to many.
27. By fairness, we should like to ensure that algorithms will not discriminate against any particular subgroup of the population. If we train a machine learning system to emulate past biased human decisions (for example in making hiring decisions), we are implicitly training the system to replicate previous human bias. In addition, developers may simply overlook the potential for difficulties, as evidenced in Google's early image recognition

system which mistook people with dark skin for gorillas. Further, there are more subtle potential problems of bias: suppose a bank is deciding whether or not to make a loan to an individual, requiring some minimum threshold level of certainty that the loan will be repaid before deciding to make the loan; if there is simply not much data available on a particular sub-population, then the algorithm will not predict a sufficiently high enough level of certainty even if the individual might otherwise represent a sound bet. Fairness in machine learning is a rapidly developing field of research so that even if we only have biased training data, still we can hope to enforce fairness constraints so that a learned system will behave in an appropriate manner.

28. Many machine learning approaches learn from a set of training data, aiming to achieve low average performance error on test data which is assumed to come from the same distribution. However, in real-world applications, we shall often require a much greater level of robustness. For example, autonomous vehicles must perform well across all weather scenarios, even those which might never have been seen previously. There are many technical challenges which must be addressed.

Question 9: In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

Ricardo Silva (Turing Fellow):

29. The power of machine learning on exploiting data as a way of avoiding explicit programming does not mean it can compensate for limitations in the data. Lack of transparency is manageable if the gaps in the data are known not to be of major importance as in, for instance, the informal use of machine translation. In fact, there is no such a thing as a complete dataset: data will contain biases and gaps that are absorbed and filled up by methods having varying degrees of transparency. The less well understood the data is as a means to achieve a particular goal, the more important transparency will be. One should not decouple transparency from data quality assessment.

Adrian Weller (Turing Fellow):

30. Transparency is often helpful, but it is not a universal good – see “Challenges for Transparency”, Weller, 2016. There are cases where it may be essential – for example, if AI is used to help decide criminal sentencing, an intelligible explanation of the reasoning should be provided to demonstrate that appropriate process was followed and to enable meaningful challenge. Transparency is sometimes a proxy for what is really required, such as reliability or fairness, and it may be more efficient to try to improve those end goals directly. Actors with misaligned interests can abuse transparency leading to a worse outcome (for example, gaming of a system). In some cases, transparency must be considered in tension with privacy requirements.

The role of the Government

Question 10: What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Brad Love (Turing Fellow):

31. The Government has a constructive role to play in the development and use of artificial intelligence as it does with other emerging technologies. However, it could be detrimental to treat AI differently than any other promising technology that is steadily improving. Policy and regulations should be informed by the realities and practicalities of this technology, rather than alarmist voices.
32. Existing laws and regulations may adequately cover AI. For example, currently, an aircraft manufacturer would be liable for a malfunctioning autopilot system that led to a loss of life. If, instead, the autopilot were "artificially intelligent", which systems developed decades ago could be considered, the same responsibilities would hold. We already have laws that cover faulty products, as well as the release of computer code (e.g., viruses) that are intended to harm the general public.
33. Specific regulation of AI would be difficult to craft as there is no firm line dividing AI from non-AI approaches. AI and machine learning approaches are continuous with standard engineering and statistical approaches. What was once considered AI, like product scanners in the supermarket, are now commonplace. Insurance firms use complex statistical models to evaluate risk and policy premiums. Should these techniques be considered AI?
34. If a guideline could be established, it would be very difficult to determine whether an AI system was in compliance. In complex AI systems, it is not always clear why a system made the decision it did, which parallels difficulties in determining the bases of performance in human experts. One active area of research in AI is trying to understand the bases of performance in AI systems, much like how Experimental Psychologists continue to examine the bases for human performance.
35. AI specific regulation could reduce innovation and competitiveness for UK industry. Consider a corporation developing an AI smoke detector that was designed to use context to be more sensitive to actual fires while avoiding false alarms. Such a system could save lives and inconvenience, but the competitive benefit in the market could be overshadowed by additional regulatory burden. Like the autopilot example above, existing laws and testing standards regulating such products should prove satisfactory. Finally, the UK is only one nation amongst many competitors. Innovation will likely take place where government is supportive and adopts a measured approach, as has proved to be the case with autonomous vehicles.

Adrian Weller (Turing Fellow):

36. AI presents tremendous opportunities for all of society. Government can help build on our nation's strengths as a world leader in research and business through: increased funding for basic research; nurturing a thriving environment for start-ups and responsible business; education; and encouraging involvement of people with diverse backgrounds. Yet AI also presents important concerns, which we should address thoughtfully with appropriate governance. We support calls for a stewardship body to consider the long term governance of AI systems, their applications and their impact on society.

6 September 2017

Mr Jaafar Almusaad and Mr Philip Bree – Written evidence (AIC0039)

SUBMISSION TO THE HOUSE OF LORDS SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE (AI)

1 Artificial Intelligence: Definition and the Current State

Artificial Intelligence (AI) can be described as computer software that attempts to learn and extract patterns from data to create decision points (i.e. condition checks). AI is fundamentally different from conventional software where programmers have to “hard code” decision points. Simple AI algorithms are ineffective in capturing complex, non-linear patterns while complex algorithms are difficult to understand and therefore, to interpret.

AI has been applied to many problems with varying degrees of success. Computer vision has seen tremendous achievement, exceeding human performance in some cases. Reasonable progress has been made in Natural Language Processing (NLP). There is, however, considerable room for development when it comes to interpreting the output of some of the widely used algorithms such as Deep Learning.

With Moore's Law (i.e. exponential increase in computing power every 18 months) continuing to be applicable, ensures access to greater computing resources, AI systems will become exponentially capable but at the same time more difficult to interpret. For instance, **Deepmind**'s Alpha Go incorporates nearly 100 hidden layers.

There is a trend to make AI available as a Service (AIaaS). This could encourage more individuals and organisations to use it, just like Cloud Computing.

3 Preparing General Public for AI

While AI is not exactly a new topic, it has only become mainstream in recent years. As such, the current generation of school teachers tends to be not sufficiently informed about AI. In order to prepare the young generation for the future, alternative means of education needs to be sought. A good start would be BBC Children and/or BBC Player Kids.

To prepare young adults to embrace AI, again the various BBC channels should take the lead so that the Mainstream Media follows. AI needs continue with well-informed coverage in simple, non-technical jargon suitable for public consumption.

It is important to demystify AI because the general public appears to have ungrounded fears that robots may take over humanity in the future (i.e. The War of the Worlds simulated news reports by Orson Welles). A trusty change

management framework can be adopted to mitigate the initial state of denial by the public and guide the change of public perception through subsequent stages: denial, resistance, exploration, acceptance and finally confidence.

7 AI Monopoly Addressed

The fierce competition between large multinational corporations, and recent multibillion acquisitions should serve as a clear indicator/warning that AI is arguably becoming a detrimental factor for business survival. Smaller players are unlikely to catch up with such fierce competition and face the risk of being left behind.

A practical approach to mitigate this risk is to “democratise” AI. This democratisation process can only be accomplished if Government is prepared to allocate sufficient public funds to support the development of open and advanced AI platforms. Government can also play a far more active role in standardising Open Data so that, amongst others, smaller companies, universities and researchers can advance their research and generate value from AI. For instance, the UK Government has already created an initiative for Open Data (data.gov.uk), which is a [positive] step in the right direction. However, much more effort is needed, especially when it comes to data standardisation. This initiative can be scaled up to include data from the private sector after the standardisation issue is addressed.

The general public need to be well informed about what data is collected, stored, processed, and also, more importantly, if data is generated about them by AI.

10 The Role of The Government

Since AI is at a very early stage, many commercial enterprises are likely to attempt to influence proposed legislation in accordance with their private interest, which could potentially conflict with their public duty of being independent from government (and have no influence on government policy). This very call for evidence on AI should not be treated as an exception. Subsequently, protective regulations will be needed by the government to protect the public interest.

Commercial enterprises need to be legally obliged to inform the public whenever AI is used in order to make decisions that are likely to have an impact on any person, whether positive or negative (dependent on the point of view). Similarly, when AI is changed in a way that could affect the outcome significantly, the public need to be clearly informed.

Authors:
Philip Bree
[LLB(Hons) PGDipLP MA(Legal Practice) LLM]

Jaafar Almusaad
MSc.(Information Technology Management)

30 August 2017

Amnesty International – Written evidence (AIC0180)

Amnesty International United Kingdom Section

www.amnesty.org.uk

Overview

1. Artificial intelligence (AI) could bring huge benefits to UK society. AI technology could help narrow the gap of inequality by widening access to advanced healthcare diagnostics and treatments; automation could take people out of dangerous and degrading work; AI systems could correct for data bias and work as anti-discrimination tools in public services. As Amnesty International Secretary General Salil Shetty stated at the 2017 AI for Good Summit, we have an incredible opportunity to use this technology for good.³⁷ However, Amnesty believes there are some key issues with AI systems that urgently need to be addressed in order to respect human rights now and protect rights in future.
2. Amnesty's chief concerns with current AI systems are that:
 - AI technology is predicted to fuel massive changes to employment in the UK, particularly through automation of jobs, which will require governmental action to protect workers' rights.
 - AI systems collecting and processing vast amounts of personal data create new threats to rights, including personal privacy rights.
 - A growing body of research demonstrates that AI systems are contributing to discrimination – for example, in policing in the US and UK.
 - A lack of transparency and accountability in current systems denies those harmed by AI-informed decisions adequate access to justice or remedy.
 - Innovation in AI is for the most part being led by corporate actors, which could lead to limiting access to AI technology to a select few in future.

Summary of recommendations

3. Amnesty International recommends that the UK government:
 - Considers and acts to protect workers' rights and the right to work where AI is predicted to heavily impact employment practices.
 - Ensures that the rights of individuals, including privacy rights, are

³⁷ <https://www.amnesty.org/en/latest/news/2017/06/artificial-intelligence-for-good/>

- strengthened and upheld in the GDPR and UK-equivalent data protection laws.
- Introduces regulation to ensure that AI systems are audited effectively and held accountable, with clear processes of responsibility.
 - Educates and informs citizens of their rights concerning privacy and data, including in automated decision-making.
 - Invests in AI developments in the public sphere to foster AI technology and solutions for the public interest.
4. Amnesty also believes that AI developments pose a huge threat to human rights in the field of conflict and policing, and calls for an international pre-emptive ban on the development, transfer, deployment and use of autonomous weapons systems.

Artificial intelligence

5. For the purpose of this paper, Amnesty defines artificial intelligence as advanced computer software and computer-powered hardware that can undertake advanced computational or physical tasks using large amounts of data to guide their operations.

Amnesty International UK

6. AI advancements will have a significant economic and social impact in the UK in the near future and Amnesty welcomes the opportunity to feed into this Inquiry.
7. Amnesty International UK is a national section of a global movement of over three million supporters, members and activists. We represent more than 600,000 supporters, activists and members in the United Kingdom. Collectively, our vision is of a world in which every person enjoys all the human rights enshrined in the Universal Declaration of Human Rights and other international human rights instruments. Amnesty's mission is to undertake research and action focused on preventing and ending grave abuses of these rights. We are independent of any government, political ideology, economic interest or religion.

Impact on society

Q3. How can the general public best be prepared for more widespread use of artificial intelligence?

8. There are two major areas where the government must place attention and invest resources in order to ensure that widespread use of AI benefits

and does not erode human rights: in employment and in personal privacy protections.

Impact of AI on employment and workers' rights

9. Advanced AI software will likely increase automation in the workplace, as systems become adept at more complex tasks. Technological advances and 'efficiency' savings will likely see machines replacing people in the workplace, as roles become part or fully automated.
10. A PricewaterhouseCoopers forecast this year estimated that up to 30% of current jobs in the UK could be automated in the next 15 years, putting over 10 million people in the UK out of work.³⁸ The Bank of England anticipates that up to 15 million jobs could be at risk of automation.³⁹
11. The UK government needs to ensure that people in the UK can access their employment rights now and in the future, including:
 - Investing in training and reskilling programmes to help those whose jobs could be at risk of automation to stay employable, considering new skills that will be in demand in a tech-driven economy.
 - Preparing for an employment landscape that is radically altered by mass unemployment and fully considering the impact on state welfare and benefits systems. This may include exploring the viability and desirability of alternative income models like Universal Basic Income.⁴⁰

Personal data – privacy and profiling risks

12. Advancements in AI come hand-in-hand with the development of vast economies of personal data – raising concerns about privacy rights. AI systems are developed and trained using extremely large datasets. They are by and large designed to hone their function through continually processing new data – the larger quantities of relevant data that the system can access, the better. (For example, AI software in healthcare diagnostics will in theory perform better over time through collecting and processing live data from a wide source of patients in order to create more accurate diagnoses).

³⁸ http://pwc.blogs.com/press_room/2017/03/up-to-30-of-existing-uk-jobs-could-be-impacted-by-automation-by-early-2030s-but-this-should-be-offse.html

³⁹ <http://www.bankofengland.co.uk/publications/Pages/speeches/2015/864.aspx>

⁴⁰ For more on the human rights case for exploring Universal Basic Income, see report by Philip Alston, UN Special Rapporteur on Extreme Poverty and Human Rights, delivered to the UN Human Rights Council in June 2017: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G17/073/27/PDF/G1707327.pdf>

13. There are numerous risks associated with networked systems storing and processing such large amounts of personal data:
- Use of advanced AI software will dramatically increase the points of personal data collection in terms of both volume and detail. For example, facial-recognition and gait recognition technologies can easily capture and process detailed personal information on a previously unforeseen scale.
 - The networking of interconnected systems – from the internet and telecoms, to systems and sensors in travel, health, logistics, traffic, electricity networks – allows the possibility for cross-referencing data that, if collected previously, used to be held in silos. Networked big data may be used to create intimate and precise personal profiles of individuals, a tactic already widely used for commercial advertising and political marketing during elections.⁴¹ AI software makes profiling on such an intimate individual level much more accessible – with the potential for companies and governments to influence people to a greater degree than ever before, using highly personalised messaging across a range of platforms.
 - Personal data is increasingly being used by systems to inform decision-making processes in all areas of our lives. There is potential for discrimination where information from one aspect of someone’s life or previous behaviour is used to inform a decision or access to a service elsewhere. For example, insurance providers may use social media data to evaluate an insurance claim without the claimant’s knowledge.⁴²
14. To ensure personal data collection and use by AI systems does not impact negatively on the rights of people in the UK, the government must:
- Ensure that the rights of individuals, including privacy rights, are strengthened and upheld in the GDPR and UK-equivalent data protection laws.
 - Invest in public education to make people more data literate and aware of their rights. This means ensuring that individuals know not only what their rights are, but how to make a complaint or seek redress where they feel their data has been misused.
 - Give greater powers to regulatory bodies that provide oversight and accountability on the use of AI and big data, particularly where AI systems could adversely affect rights.
 - Ensure adequate regulation of private companies, including, for example, by mandating independent audits of AI systems where their use case means they have the potential to significantly impact human

⁴¹ <http://www.bbc.co.uk/news/uk-39171324>

⁴² Car insurance company Admiral last year attempted to use Facebook data to glean information that would inform insurance decisions: <https://www.theverge.com/2016/11/2/13496316/facebook-blocks-car-insurer-from-using-user-data-to-set-insurance-rate>

rights.

- Ensure that AI systems in public service use are designed in a manner compatible with human rights standards, such as being non-discriminatory and providing means to pursue effective remedy.
- Require that all AI systems used in public services and other services that directly impact on human rights are clearly identified as AI systems. Companies and public bodies should always disclose when such a system is used to deliver services or make decisions that impact people's rights.

Q4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

15. Amnesty has three main concerns about how developments in AI may negatively impact upon individual's rights:

- AI systems are capable of perpetuating or facilitating discrimination.
- A lack of transparency and accountability in current systems denies those who are harmed by AI-informed decisions adequate access to justice or remedy.
- AI systems that collect and network vast amounts of personal data threaten rights, including personal privacy rights.

AI systems may perpetuate or facilitate discrimination.

16. The adoption of AI and data-driven processes to aid governance and decision-making across many sectors of society has the potential to facilitate discrimination if proper oversights are not put in place.

17. Getting approval for a loan or mortgage or purchasing health or home insurance in the future will increasingly be determined by personal data run through an unaccountable algorithm. As argued by data scientist Cathy O'Neil, such systems are capable of reinforcing and entrenching existing discrimination based on profile data such as income, home address, ethnicity, gender or religion.⁴³ At the same time, the algorithm's decision is frequently beyond scrutiny. An individual who is charged a higher premium for their insurance or denied a mortgage has no means of challenging this decision and interrogating the data upon which it is based.

18. There is already worrying evidence that the use of AI and big data in

⁴³ For example, see Cathy O'Neil's 'Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy' (2016)

policing can perpetuate discrimination and identity bias. As predictive policing systems advance rapidly and are deployed across the law enforcement and security spheres, there is an urgent need to put safeguards in place to minimise the risk of human rights abuses and guarantee accountability when errors are made. Scrutiny of such systems and how they work as 'decision support' tools in the police is difficult, given that these systems are usually proprietary.

19. One research study from the Human Rights Data and Analysis Group (HRDAG) in the US developed a replica of a predictive policing algorithmic programme that is used by police forces in numerous US states, and ran it as a simulation on crime data in Oakland. They concluded that the programme reinforced existing racial discrimination within the police. This was because the system was built using already biased data that recorded higher crime rates in parts of the city with a higher concentration of black residents. The algorithm therefore predicted more crime in those areas, dispatching more frontline police officers, who unsurprisingly made more arrests. The new data was fed back into the algorithm, reinforcing its decision-making process and creating a pernicious feedback loop that would contribute to over-policing of black neighbourhoods in Oakland.
20. In the UK, Kent and Essex police forces are piloting the same software replicated in the HRDAG study.⁴⁴ Given that there has been no public evaluation of this pilot or open audit of the software, it is not clear whether deployment of this system has carefully considered and mitigated against bias.
21. Other police forces in the UK are testing the use of automated systems in policing. The Metropolitan Police Service (MPS) police deployed automated facial recognition software at Notting Hill Carnival for the second year running.⁴⁵ Research has shown that facial recognition algorithms can carry racial biases – which could inadvertently lead to discriminatory policing. The FBI's software misidentifies people almost 15 per cent of the time – and is more likely to fail with black people and women. If the MPS's software has similar flaws, using it at events like Notting Hill Carnival raises serious human rights risks.⁴⁶
22. Furthermore, in 2014, the MPS announced it would introduce an automated system to assign risk scores to individual suspected of being

⁴⁴<https://www.whatdotheyknow.com/request/181341/response/454199/attach/3/13%2010%20888%20Appendix.pdf>

⁴⁵ <https://www.theguardian.com/uk-news/2017/aug/05/met-police-facial-recognition-software-notting-hill-carnival>

⁴⁶ <https://www.liberty-human-rights.org.uk/news/press-releases-and-statements/undemocratic-unlawful-and-discriminatory-civil-liberties-and-race>

'gang members' in London.⁴⁷ Reportedly, the pilot used data gleaned from social media along with police crime reports to generate offending risk scores for all individuals associated with London gangs.

23. Amnesty International's ongoing research into the MPS's Gangs Databases suggests that the current manual system used by the police to flag individuals as 'gang associated' is arbitrary, lacks adequate oversight and contributes to the overrepresentation of BAME young people in the criminal justice system. In this context, the introduction of automated risk-scoring on top of an already deeply flawed data collection policy with no effective oversight and safeguards in place raises significant human rights concerns.

A lack of transparency and accountability in current systems denies those who are harmed by AI-informed decisions adequate access to justice or remedy.

24. The inability to scrutinise the workings of all current deep learning systems (the 'black box phenomenon') creates a huge problem with trusting algorithmically-generated decisions. Where AI systems deny someone their rights, understanding the steps taken to deliver that decision is crucial to deliver remedy and justice.

25. Provisions for accountability need to be considered before AI systems become widespread – practically, this may occur at multiple points, including in developing software, using training data responsibly, executing decisions. To what extent will any automated decision be able to be 'overridden', and by whom?

26. Restricting the use of deep learning systems in some cases may be required, where such systems make decisions that directly impact individual rights. The UK government should encourage the development of explainable AI systems, which would be more transparent and allow for effective remedies.⁴⁸

27. Systems need transparency, good governance (including scrutiny of systems and data for potential bias), and accountability measures in place before they are rolled out into public use – especially where AI systems play a decisive and influential role in public services (policing, social care, welfare, state healthcare).

⁴⁷ <http://www.bbc.co.uk/news/technology-29824854>

⁴⁸ For example, a draft bill before New York City council advocates for transparency for all systems where algorithms are generating decisions in government services: <https://www.nytimes.com/2017/08/24/nyregion/showing-the-algorithms-behind-new-york-city-services.html>

AI systems that collect and network vast amounts of personal data threaten rights, including personal privacy rights.

28. The right to privacy is hugely significant and yet widely abused by states (through government mass surveillance programmes) with advances in technology, which many governments, including the UK, have ultimately taken advantage of to access and store private information on an unprecedented scale. The 2016 Investigatory Powers Act granted the UK government extraordinarily wide powers to view, obtain and store personal data.
29. As outlined earlier, the proliferation of AI systems creates the possibility for system owners to collect detailed and intimate personal information at an individual level. There is a risk that corporate actors, states or individuals with access to vast amounts of personal data could hold huge sway over those whose personal data they can access and therefore influence.

Industry

Q7. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

30. Government and civil society have struggled to keep up with the myriad of challenges to privacy and freedom of expression posed by developments in internet technologies: laws and public policies are still catching up with technologies that have been in wide use for years, if not decades. At the same time, there is a tension for policy-makers between the imperative to get to grips with and regulate the development and use of AI systems, and the tempting appeal of these systems – which promise to ‘modernize’ and ‘increase efficiency’ across the public sector, while reducing cost. The overwhelming majority of AI systems are developed by private technology companies – systems which governments then may purchase to use in public services. As the uses of powerful AI technologies start to permeate all aspects of life, it is crucial that civil society and governments do not lag behind in responding to AI developments as they did with the development of the internet.
31. Amnesty is concerned that proprietary AI systems built by private actors will be in widespread use, including across the public sector, before human rights risks have been fully considered and addressed. This presents a major barrier to ensuring transparency and accountability of such systems.
32. Furthermore, we are witnessing a trend whereby AI research and development is being driven by the corporate sector, with companies quick to hire AI scientists and experts who have previously worked in the public

sector and academia.

33. Amnesty's concern is that AI technologies that could bring positive developments to many, for example in healthcare, will be restricted by the commercial imperatives of companies, and that the benefits of innovation could be restricted by aggressive intellectual property practices.
34. Amnesty recommends that the UK government considers creating a well-resourced public AI initiative that significantly invests in the development of AI technologies and solutions that are in the public interest, and facilitates widespread use of beneficial AI technology. Otherwise, major technology companies will continue to almost exclusively control the innovative technology that could end up being used to deliver key public services. A publicly-funded AI initiative could provide open source access to AI technologies, or free/low cost licensing, for public good purposes. While the private sector can play an important role in developing AI uses for the public good, it should not exclusively shape it.

Ethics

Q8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

35. Amnesty's chief concerns with current AI systems are that:
- A growing body of research demonstrates that AI systems are contributing to discrimination, for example in policing in the US and UK.
 - A lack of transparency and accountability in current systems denies those harmed by AI-informed decisions adequate access to justice or remedy.
 - AI systems collecting and processing vast amounts of personal data create new threats to rights, including personal privacy rights.
36. Amnesty International recommends that the UK government:
- Ensures that the rights of individuals, including privacy rights, are strengthened and upheld in the GDPR and UK-equivalent data protection laws.
 - Creates and upholds adequate regulation of private companies, including, for example, by mandating independent audits of AI systems where their use cases means they can potentially have a significant impact on human rights.
 - Guarantees that AI systems used in public services are designed in a manner compatible with human rights standards, such as being non-discriminatory and providing the possibility of effective remedy.
 - Considers restricting the use of AI systems that can't be interrogated

(such as deep learning), where those systems generate decisions that affect individual or a groups' enjoyment of their human rights.

- Gives greater powers to regulatory bodies that provide oversight and accountability on the use of AI and big data in the delivery of services, including strengthening the mandate of the Information Commissioner's Office (ICO).
- Educates and informs citizens of their rights concerning privacy and data, including in automated decision-making.

38. Amnesty also believes that AI developments pose a very serious threat to human rights in the field of conflict and policing, and calls for an international pre-emptive ban on the development, transfer, deployment and use of autonomous weapons systems.

Q9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

39. The black box problem is most acute in applications of AI that have the potential to impact on individual rights. Many commercial and non-commercial AI applications may not require the same threshold of accountability, as their impact on individual rights is either remote or negligible.

40. It is vital that AI systems are not rolled out in areas of public life where they could discriminate or generate otherwise unfair decisions without the ability for interrogation and accountability.

41. As stated above, where there is potential adverse consequences for individual rights, there must be higher transparency standards applied, with obligations both on the developers of the AI and the institutions using the AI system. This includes:

- Detecting for and correcting for bias in design of the AI and in the data used.
- Effective mechanisms to guarantee transparency and accountability in use, including regular audits to check for discriminatory decisions and access to remedy when individuals are harmed.
- Not using AI where there is a risk of harm and no effective means of accountability.

Transparency problems with Autonomous Weapons Systems

42. One particular circumstance where a lack of transparency means that use is never acceptable is in the case of Autonomous Weapons Systems (AWS).

43. The development, deployment and use of AWS raises important issues related to transparency and accountability for human rights violations and individual criminal responsibility. Use of AWS would pose serious challenges to bringing accountability for crimes under international law. Under international human rights law, states have an obligation to investigate allegations of human rights violations and bring the perpetrators to justice as part of the right to an effective remedy – a right which is applicable at all times.
44. In the case of lethal and less-lethal AWS, it is not possible to bring a machine to justice and no criminal sanctions could be levelled against it. However, actors involved in the programming, manufacture and deployment of AWS, as well as superior officers and political leaders, should be accountable for how AWS are used. But the nature of AWS is such that it would be impossible to foresee or programme how an AWS will react in every given circumstance, given the countless situations it may face.
45. Furthermore, without effective human oversight, superior officers would not be in a position to prevent an AWS from committing unlawful acts, nor would they be able to reprimand it for misconduct. AWS, by their very nature, are autonomous agents that have no individual accountability. Deploying them in combat or for the use of force in civilian environments would be a perilous step for humanity, taking away one of the strongest deterrents against the unlawful use of violence.

The role of the Government

Q10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

46. The UK government should:
- Creates and upholds adequate regulation of private companies, including, for example, by mandating independent audits of AI systems where their use cases mean they can potentially have a significant impact on human rights.
 - Ensure that AI systems in public service use are designed in a manner compatible with human rights standards, such as being non-discriminatory and providing means to pursue effective remedy.
 - Invest in AI development in the public sphere to ensure development of AI technology and solutions for the public interest, and that it does not solely follow the commercial interests of private companies.
 - Ensures that the rights of individuals, including privacy rights, are

strengthened and upheld in the GDPR and UK-equivalent data protection laws.

- Educates and informs citizens of their rights concerning privacy and data, including in automated decision-making.
- Consider restricting the use of AI systems that can't be interrogated (such as deep learning), in use cases where those systems make automated decisions that affect an individual or a groups' enjoyment of their human rights.
- Advocate for a pre-emptive international ban on the development, transfer, deployment and use of Autonomous Weapons Systems.⁴⁹

Learning from others

Q11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

47. Amnesty advocates for a cooperative approach to understanding advances in AI and mitigating against risks. Amnesty recommends that the government continues to consult civil society organisations as well as technology specialists, to ensure that AI developments protects and promotes human rights.

6 September 2017

⁴⁹ Amnesty International urges the UK government to engage in a comprehensive debate around the multiple challenges posed by AWS in order to develop and articulate a national policy on AWS (including less-lethal AWS) that takes full account of the state's obligations to respect and ensure international human rights law and international humanitarian law. This must be done in consultation with a broad range of stakeholders, including by meaningful and substantive engagement with non-governmental organizations and relevant experts, including AI and robotics experts and industry leaders.

Dr Sally Applin – Written evidence (AIC0172)

- 1) The oncoming onslaught of Artificial Intelligence (AI) is not something that will happen to humanity, but rather something that we ourselves will construct, shape, and enable in the world. Some of us may have more power than others in its implementation and deployment of AI. It is for this reason that is astute for those shaping the governance of our future to both gather data and understanding of concerns, and to take action to protect not only their constituents, but broader humanity and global society—for as we all now realise, digital networks and digital automation is broadly reaching and the smallest digital intent can have unforeseen global repercussions.
- 2) There are two points that I would like to personally contribute to for this call, the first being Human Agency and its preservation, and the second being that of Social and Cultural awareness when automating decisions that will impact ethics. Human agency is our capability to make choices and decisions from the options that unfold before us at each point in time. As we move through the world, and as our circumstances change, so do the options from which we may choose to make any given decision. When these are automated, and in the case of AI, severely estimated and automated, the results can restrict human freedom and movement—in any class of society. Furthermore, because these decisions are automated, the cultural and social aspects of each individual as well as our cultural groups, does not become considered. This can undermine peoples' agency as well as their identity. I refer to ethnicity and agency within a country's national identity as part of a discussion on ethics, values, and customs within a culture, as well as individual agency and cultural expression within that context. An AI from Michigan in an autonomous vehicle with embedded ethics would suggest one type of cultural values, which may be out of place in Great Britain, where people express their cultural values in different types of vehicular ethical behaviour. What does it mean to automate cultural choices and expressions in one area, and deploy those to other locales? (See Applin 2017: <http://ieeexplore.ieee.org/document/7948873/>)
- 3) Automation currently employs constructed and estimated logic via algorithms to offer choices to people in a computerised context. At the present, the choices on offer within these systems are constrained to the logic of the person or persons programming these algorithms and developing that AI logic. These programs are created both by people of a specific gender for the most part (males), in particular kinds of industries and research groups (computer and technology), in specific geographic locales (Silicon Valley and other tech centres), and contain within them particular "baked-in" biases and assumptions based on the limitations and viewpoints of those creating them. As such, out of the gate, these efforts do not represent society writ large nor the individuals within them in any global context. This is worrying. We are

already seeing examples of these processes not taking into consideration children, women, minorities in terms of even basic hiring talent to create AI. As such, how can these algorithms, this AI, at the most basic level, be representative for any type of population other than its own creators?

- 4) The impact on society of the digital revolution has had a profound global societal impact and the issues that we have seen with Google and Facebook bumping up against privacy laws and regulations in Europe are a direct result of this cultural mismatch and lack of awareness of other ways of living and life. Thus, one important and critical step for government would be to mandate that teams developing AI include research scientists and contributors from multiple cultures, social classes, ethnicities, and genders.
- 5) If this does not happen, the representative power and advantage is distilled into a very small group of people, designing a system mostly for themselves, with the power and capabilities to extract habits, data, and behaviours from others, all concentrated within the power of technology companies. This is a problem that is ongoing. Google and Facebook have more data (and more relevant data) on citizens than most governments.
- 6) If the companies building this future do not include most of humanity—how could the AI they produce be fair, representative, and appropriate for societies?
- 7) Additionally, the government should include Social Scientists, particularly anthropologists on a panel or task force as these debates move forward. Anthropologists specialise in understanding groups, and group cultural behaviour, and there are some of us who have training in technology and technology development. I have spent the summer in a Silicon Valley multinational corporation's AI group, observing how AI decisions are made and deployed. My conclusion from this experience is that many of my concerns regarding balance in AI are founded, and will need addressing as we move to build an automated and intelligent future.
- 8) The public should be made aware that their choices are being limited by AI, that their cultures and genders are not being fully considered by AI, and that if they want to have a choice, true agency and choice, equivalent or better to what they have now, that they must understand how critical it is that AI development teams be balanced and representative, and that all of us are included in the shaping of our future.
- 9) Our papers address this and can be found at <http://www.posr.org/wiki/publications> - specifically, Applin and Fischer (2015): "New Technologies and Mixed-Use Convergence: How Humans and Algorithms are Adapting to Each Other" (<http://ieeexplore.ieee.org/document/7439436/>), where we explore cultural

relations to automaton in the context of human agency.

Sally Applin, Ph.D.
University of Kent, Canterbury, UK
School of Anthropology and Conservation
Centre for Social Anthropology and Computing

.....
Associate Editor, IEEE Consumer Electronics Magazine
Member, IoT Council
Board Member: The Edward H. and Rosamond B. Spicer Foundation

6 September 2017

Arm – Written evidence (AIC0083)

The pace of technological change

- 1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?**

Much of the following is drawn from a report which Arm commissioned earlier this year from Northstar Research Partners Ltd based in London. (Full text available from contact below.)

Predicting the future of technology is never easy. There is an observation attributed to Bill Gates about always overestimating what change will happen in 2 years and underestimating what will happen in 10 years probably applies. Predicting tech much beyond 7 years is a challenge.

That said, it is worth noting that Sundar Pichai, CEO of Google, wrote last year that “the last 10 years have been about building a world that is mobile-first. In the next 10 years, we will shift to a world that is AI-first.”

Artificial Intelligence (AI) has been talked about for decades. It is a label used to cover a variety of technologies including machine learning and robotics. Some would say AI is mostly machine learning. Others might argue that AI goes a step further, requiring machines to ‘think’ for themselves in ways similar to human beings. Robots are probably machine learning/AI devices which move.

In recent years it is clear that AI is beginning to become a reality for consumers thanks to the rise of autonomous vehicles and personal assistants such as Siri and Google Now.

And Tech Sector investment in AI is growing very fast. Forrester Research has predicted a “greater than 300% increase in investment in artificial intelligence in 2017”, compared with 2016, a testament to the sector’s rapid global growth.

Experts tend to agree that AI technology is at a tipping point and could have a profound impact on the world in the near and long-term future.

- 2. Is the current level of excitement which surrounds artificial intelligence warranted?**

Yes. It is a once in a generation changes not only in how computing gets done, but also in its potential impact on a wide range of human activities.

According to Northstar’s global survey, three quarters of consumers expect AI to feature heavily in their lives by 2022. Just over a third of people now think AI is already having a notable impact on their daily lives. But most people globally still seem to believe the AI evolution is only just gathering pace.

Impact on society

- 3. How can the general public best be prepared for more widespread use of artificial intelligence? In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.**

Several studies have been conducted into the likely impact of AI on jobs. Their conclusions vary. But we must act on the assumption that even if the precise outcome is hard to predict, people will be worried about the risk of job losses and this anxiety will need to be addressed in public policy.

The anxiety will not be confined to low skilled jobs. For example a number of legal/paralegal jobs may be threatened by Machine Learning replacing a lot of drafting and reading of other people’s drafting. The Northstar survey findings suggest consumers think that manufacturing and banking jobs are most under threat, while jobs in science and healthcare are the safest.

But even in these sectors there will be challenges: some jobs currently considered ‘highly-skilled’ may actually turn out to be ‘highly-experienced’ (think medical diagnosis), and thus be good targets for replacement by AI.

It is very difficult to predict exactly how AI will have an impact on jobs.

One possibility is that a mixed workforce of AI machines and humans may be the most likely outcome in most industries.

AI will also create new job opportunities. The recent rise of internet-based service models points to the positive impact large scale transformation can have on employment.

The challenge of preparing a workforce for the new jobs is well known: how can we devise education systems which allow people to cope through retraining, acquiring key skills etc? How can we ensure that there will be decent livelihoods for all types of people?

As the World Economic Forum has said: “65% of children entering primary /elementary school today will ultimately end up working in completely new job types that don’t yet exist”.

Addressing this challenge requires all those involved in UK education including universities and FE colleges to consider how they can adjust their teaching courses and even their methods. This is not simply a matter of training schoolchildren in STEM, although that is important. It is also about offering opportunities for retraining and upskilling later in life. FE colleges need to find the niche areas (like user interface design) which will attract broad interest from those looking to find work in the digital era.

According to our survey concerns about jobs were the biggest worry for consumers.

But there were also concerns about the impact of AI for increased data sharing, and data security and protection.

85% of global consumers were concerned about the security of AI. US and European consumers were more likely to worry about the “reliability” of AI machines, reflecting a lower positivity generally. In Asia, reliability was less of a concern, but there was genuine apprehension about AI machines becoming more intelligent than humans.

The benefits of AI will derive primarily from machines being able to learn quickly from huge amounts of data. Accordingly the need to secure the data and to promote handling of it which protects privacy, is paramount.

Arguably this is an extension of the sort of data protection arrangements which will be needed to enable the success of the other emerging digital technologies (like the Internet of Things). Efforts are underway to look at how Governments and Industry can help drive up awareness of the need for better data security and protection in these areas, including the possibility of Trust Labels, Codes of Conduct (see the work of the IoT Security Foundation) etc.

Other issues which will need to be addressed include liability: what happens when a genuine AI machine makes a decision which results in harm? In such cases unravelling the machine’s thought processes may not be straightforward.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

AI may have the potential to truly transform society for the better. For example, AI may enable advanced scientific and more complex medical research, where the ability to assess vast quantities of data in superfast time and use a computer brain to make instant decisions can bring benefit.

The Northstar Survey indicates that consumers believe that AI technology and increased automation will help rather than hurt society. Well over half of consumers predict a better societal future, versus a fifth who expect things to get worse. Asian consumers see the brightest prospects, with around three quarters predicting positive change. The picture is more balanced in Europe, where there was greater scepticism on the likely impact of AI.

At company level To date, it seems that those companies that own high-quality datasets will be best placed initially to benefit commercially from the advance of AI. This is mostly in the hands of a small number of multinational and national companies.

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

Yes. This is something that has the capacity to affect millions/billions of people’s lives. There are clearly mixed emotions around the impact of AI. This balance may shift if on the one hand it becomes clear that AI is applied in ways that help preserve human health and enhance people’s quality of life, and if fears about job losses are addressed.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

7. How can the data-based monopolies of some large corporations, and the ‘winner takes-all’ economies associated with them, be

addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Ethics

- 8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.**

See points on privacy etc above. It has sometimes been said that we need a culture of 'ethical by design' for AI, in the same way that we are trying to promote 'secure by design' for other connected devices.

- 9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?**

The role of the Government

- 10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?**

The Government will have a role in helping to build citizen confidence in AI in some of the ways mentioned above. It is too soon to talk of regulation. If we were to go for an 'ethical by design' approach, it would be normal to start with a Code of Practice or key principles.

Learning from others

- 11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?**

Arm Stephen Pattison

VP Public Affairs

4 September 2017

Article 19 – Written evidence (AIC0129)

1. ARTICLE 19: Global Campaign for Free Expression (ARTICLE 19) welcomes the opportunity to respond to this inquiry by the House of Lords Select Committee on Artificial Intelligence (AI). ARTICLE 19 is a global human rights organisation that works around the world to protect and promote the right to freedom of expression and information ('freedom of expression'). Established in 1987, with its international head office in London, ARTICLE 19 monitors threats to freedom of expression in different regions of the world, and develops long-term strategies to address them. We advocate for the implementation of the highest standards of freedom of expression, nationally and globally.
2. Since 2014, ARTICLE 19 has pioneered efforts in technical communities to bridge existing knowledge gaps on human rights and their relevance to internet infrastructure. Our efforts have been geared towards integrating human rights into foundational documents at the Internet Corporation for Assigned Names and Numbers (ICANN),⁵⁰ the Internet Engineering Task Force (IETF)⁵¹ and the Institute for Electrical and Electronics Engineers (IEEE).⁵² At the IEEE specifically, ARTICLE 19 has taken an active part in the Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems.⁵³ In December 2016, we also published a policy brief on algorithms and automated decision-making in the context of crime prevention.⁵⁴
3. The study and development of AI is over half a century old, with the term being coined in 1956.⁵⁵ The current momentum in this field is enabled by the availability of large amounts of data, computational power that is both affordable and widely accessible, the continued development of statistical methods and the mainstream adaptation of technology. Hence, ARTICLE 19 believes in the need to critically evaluate the impact of AI and automated decision making systems (AS) on human rights, and the various ways in which these technologies embed values and bias, thereby strengthening or

⁵⁰ See, for example ARTICLE 19, ICANN's Corporate Responsibility to respect Human Rights, October 2015; available at <http://bit.ly/1KgkV5n>.

⁵¹ See, for example ARTICLE 19, Internet Engineering Task Force discusses human rights in plenary meeting for the first time in its history, April 2017; available at <http://bit.ly/2wwz037>.

⁵² See, for example ARTICLE 19, A New Frontier: Ethics, Artificial Intelligence and the Institute of Electrical and Electronics Engineers (IEEE), December 2016; available at <http://bit.ly/2wwBEps>.

⁵³ See, The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, Ethically Aligned Design, available at <http://bit.ly/2pIPsMc>.

⁵⁴ ARTICLE 19, Algorithms and automated decision-making in the context of crime prevention, 2 December 2016; available at <http://bit.ly/2gSnG9W>.

⁵⁵ Stuart J Russel & Peter Norvig, *Artificial Intelligence – A Modern Approach*, Englewood Cliffs, NJ: Prentice Hall, 1995: 27.

sometimes hindering the exercise of these rights, particularly freedom of expression. The role of industry, governments, and individual developers must be grounded, at the very minimum, in existing standards of corporate responsibility and international standards on human rights. Given our mandate, this submission focuses on the issues most directly connected with defending freedom of expression and information.

4. **Terminology:** At the outset, ARTICLE 19 notes that the terminology around AI varies and can encompass different concepts, in particular:

- “Algorithm” can refer to any computer code that carries out some set of instructions, and is essential to the way computers process data.⁵⁶ They are “encoded procedures for transforming input data into desired output, based on specific calculations”.⁵⁷
- Automatic decision making execution “generally involves large scale collection of data by various sensors, data processing by algorithms and subsequently, automatic performance.”⁵⁸ It is an efficient means to manage, organize and analyse large amounts of data and then to structure decision-making accordingly.⁵⁹
- Artificial Narrow Intelligence is the ability of machines to approximate human intelligence in a *deliberate domain*;
- Artificial General Intelligence, also commonly referred to as “Singularity”, is understood as the ability of machines to exhibit all aspects of human intelligence.

In this submission, we refer to AI only in terms of “Artificial Narrow Intelligence,” as popular perceptions of “Artificial General Intelligence”, is still, at the very least, decades away,⁶⁰ if not entirely implausible.⁶¹

The pace of technological change

5. Machine learning algorithms⁶² increasingly influence the ways in which we interact with our environments, with applications in critical sectors. For

⁵⁶ Centre for Internet and Human Rights, “The Ethics of Algorithms: from radical content to self-driving cars – final draft background paper” GCCS 2015, available at <http://bit.ly/1D7IqTx>.

⁵⁷ Gillespie T. “The relevance of algorithms” in Gillespie T, Boczkowski P., and Foot K., “Media technologies: essays on communication, materiality and society”, 2014, Cambridge MA:MIT Press (p.167).

⁵⁸ M Perel et al, “Accountability in algorithmic copyright enforcement” 2016, Stanford Technology Law Review, Forthcoming.

⁵⁹ *Ibid.*

⁶⁰Rupert Goodwins, “Debunking the biggest myths about AI”, Ars Technica, 21 December 2012; available at <http://bit.ly/2f2eqhi>.

⁶¹ Luciano Floridi, “Should we be afraid of AI?”, Aeon, 9 May 2016, <http://bit.ly/1q8UXOz>.

⁶² Machine learning is most successful subset of AI techniques, which enables an algorithm to learn from a provided dataset using statistical methods.

example, AI currently determines the information we consume online through ranking and filtering online content, most visibly on social media platforms like Facebook, YouTube, and Twitter.⁶³ AI is increasingly used for predictive policing,⁶⁴ countering violent extremism,⁶⁵ and removal of child sex abuse images or video.⁶⁶ Courts in the United States use AI to determine the risk assessments of defendants in criminal sentencing.⁶⁷ This is a trend that is also ready for deployment in the United Kingdom.⁶⁸ Machine learning algorithms also find application in the financial sector and are used to determine the eligibility of individuals for loans and mortgages based on credit scoring,⁶⁹ and corporate bond trading.⁷⁰ Algorithms are also increasingly used for network management of critical infrastructure, from the electrical grid⁷¹ to Internet routing.⁷² While the use of AI may increase efficiency of these various processes in the future, its success so far has been limited and its use often controversial.

6. ARTICLE 19 believes that at present, there is limited understanding of the ethical and legal implications of the training, development, and control of AI systems. Machine learning currently trains algorithms on datasets with a definition of “success”, i.e. a definition of what the machine must look for, what features of the data it must train on. The choice of dataset, and the definition of success ultimately shape our interaction with these technologies. For example, some applications of AI have shown that they can exacerbate the problem of discrimination by excluding minority groups from services,⁷³ products,⁷⁴ or embedding biases against marginalised populations.⁷⁵ AI can hamper freedom of expression by unduly flagging legitimate content for

⁶³ See, for example, Casey Newton, “How Youtube perfected the feed”, The Verge, 30 August 2017, <http://bit.ly/2vOBbee>.

⁶⁴ Haley Dunning, “Predictive policing gets a boost from 3m grant”, Imperial College London, 21 March 2017, <http://bit.ly/2ncWHZh>.

⁶⁵ Matt Burgess, “Google’s using a combination of AI and humans to remove extremist videos from YouTube”, WIRED UK, 19 June 2017, <http://bit.ly/2gFBC54>.

⁶⁶ “Toddler hand inspired AI sex abuse tool”, BBC, 1 December 2016, <http://bbc.in/2eDXWLO>.

⁶⁷ Julia Angwin et al, “Machine Bias”, ProPublica, 23 May 2016, <http://bit.ly/2f2eP3w>.

⁶⁸ Nick Statt, “AI driven policing has arrived”, The Verge, 10 May 2017, <http://bit.ly/2pnXN6V>.

⁶⁹ Nanette Byrnes, “An AI-fueled credit formula might help you get a loan”, MIT Technology Review, 14 February, 2017, <http://bit.ly/2lJBzt>.

⁷⁰ “Goldman expands algorithmic bond trading”, Financial Times, 16 August 2016, <http://on.ft.com/2x2tt1S>.

⁷¹ “How Artificial Intelligence is shaping the future of energy”, Open Energi, 9 February 2017, <http://www.openenergi.com/artificial-intelligence-future-energy/>.

⁷² For example, see Hao Bai, “A Survey of AI for Network Routing Problems”, <http://bit.ly/2eK3WTP>.

⁷³ Jonathan Vanian, “Amazon bows to pressure to bring same-day delivery to poor areas”, Fortune, 6 May 2016, <http://fortune.com/2016/05/06/amazon-ignore-poor-neighborhoods-deliveries/>.

⁷⁴ Will Knight, “AI Programs are learning to exclude some African-American voices”, MIT Technology Review, 16 August 2017, <http://bit.ly/2wS2W7n>.

⁷⁵ See footnote 18, above.

takedown.⁷⁶ Facial recognition AI can increasingly identify people in a crowd,⁷⁷ undermining the right to privacy and anonymity in the public sphere. This generally indicates failures in making sound legal and ethical choices at the point of training data and defining success for the algorithm. However, this only partially explains questionable outputs. There is quite simply little understanding– and therefore accountability – at present of how machines produce outputs.

7. Although the development of AI is not new, the digital environment will make it more enabling in the future, with greater volumes of data, computational power, and advances in statistical methods. Looking ahead, there is a strong tendency to implement AI across the board, making its potential even more pronounced. However, the need to think carefully through where, how, and whether AI should be implemented is an important one.⁷⁸ For example, AI may not be appropriate for tasks that require an understanding of context and judicial determination. A worrying trend is that this increased capability is not accompanied by an increase in scrutability, i.e the ability to not only see, but also to understand and investigate decisions made by, or on the basis of AI.
8. The excitement currently surrounding AI lacks clarity. There is a tendency to conflate machines that use powerful statistical and probabilistic methods to solve problems, with machines that exhibit human intelligence across domains. At present, AI can surpass human understanding within *narrow, deliberate* domains that these technologies are trained in, as was demonstrated by AlphaGo in May 2017.⁷⁹ However, we are still far away from building machines that can truly ‘think’. Even the most complex AI currently cannot begin learning without direction, a human must still guide the machine and train it on what to look for, which also involves AI amplifying the preferences and values of the trainer. While the potential of AI is indeed exciting, the current hype is largely *misinformed* by popular coverage of developments in the area. Excitement surrounding the ‘Singularity’ focuses on fictional threats while ignoring the more urgent and immediate considerations for AI.⁸⁰

Ethics

⁷⁶Julia Carrie Wong, “Mark Zuckerberg accused of abusing power after Facebook deletes Napalm girl post”, 9 September 2016, The Guardian, <http://bit.ly/2c2e0GI>.

⁷⁷ Shaun Walker, “Face recognition app taking Russia by storm may bring end to public anonymity”, 17 May 2016, The Guardian, <http://bit.ly/23VMZpb>.

⁷⁸ Ryan Calo, ‘Artificial Intelligence Policy: A Roadmap’, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3015350.

⁷⁹ “AlphaGo beats planet’s best human Go Player”, Tech Crunch, 25 May 2017, <http://tcrn.ch/2rk3Mue>.

⁸⁰ Caroline Sindors, “Dear Elon - Forget killer robots, here’s what you should really worry about”, Fast Code Design, <http://bit.ly/2wbnisu>.

9. As AI is increasingly deployed in various sectors, ARTICLE 19 considers that there is a need for a shared ethical framework within which these algorithms can function. **The development and use of AI must be subject to the minimum requirement of respecting, promoting, and protecting international human rights standards.** This would at the very least, ensure a minimum level of fairness and accountability in these processes. It is only through this minimum standard that legal-ethical considerations like fairness, accountability can be reached.
10. At the development stage, AI is embedded with values, i.e the model is made to optimise for some specific attributes, for a specific outcome, determined by the developers of these systems. The ethical implications here are the choice of the dataset used, the prioritisation of various attributes, and the safeguards put in place to ensure the promotion of fairness, scrutability, inclusivity, and accountability. For instance, the implications of a non-inclusive training dataset were exposed in 2015, when Google’s photo app was found to be tagging black people as gorillas.⁸¹ Similarly, the issue of identifying attributes and gender discrimination in machine learning was brought to test by a 2015 study at Carnegie Mellon University which found that Google advertisements for high paying jobs were more likely to be shown to men, as opposed to women.⁸²
11. At the stage of implementation, ARTICLE 19 finds that the *manner* in which AI is used gives rise to both ethical concerns and concerns for the protection of human rights, **particularly freedom of expression.** AI enables censorship in the form of content removal, prioritization, filtering, and blocking algorithms. The detection and removal of content relating to online extremism and child sex abuse images through ‘hashes’ in the UK and the US relies on AI, but at the same time risks over blocking and operates without judicial oversight, thus setting dangerous precedent. Similarly, YouTube’s content removal algorithm, ContentID, asymmetrically privileges content owners over content creators, even in the case of legitimate speech.⁸³ Presently, AI is very poor at understanding context,⁸⁴ but is made to carry out tasks that require it to do so, which means that sometimes these technologies block or enable the removal of legitimate content, evidenced

⁸¹ “Google photos identified black people as gorillas but racist software isn’t new”, Splinter News, 1 July 2015, <http://bit.ly/2gFs767>.

⁸² “Questioning the Fairness of Targeting ads online”, CMU, 7 July 2015, <http://bit.ly/2w5Mwa3>. Also, see <https://www.theladders.com/p/26101/ai-screen-candidates-hirevue> as critique on the objectivity of AI.

⁸³ <https://www.digitalmusicnews.com/2016/02/29/youtube-alters-response-to-takedown-complaints/>.

⁸⁴ <https://www.eff.org/files/AI-progress-metrics.html#Reading-Comprehension>

most recently by Facebook’s takedown of a Pulitzer prize-winning Vietnam War photograph of a naked girl.⁸⁵

12. Potential ethical implications of AI are difficult to determine because usually, the use, underlying values and problems within these systems become apparent only when a harm arises.⁸⁶ Resolving negative implications needs to start with ensuring that the ethical framework within which AI functions has a strong grounding in international human rights standards as a minimum level of protection.
13. A relative lack of transparency, i.e black boxing, or making the logic or data being used by an AI system selectively available, is acceptable **only** where absolute transparency would involve the violation of fundamental human rights, particularly the disclosure of personal or sensitive data of individuals.⁸⁷
14. While transparency in AI systems is desirable, it is not in and of itself sufficient to hold algorithms accountable.⁸⁸ It is important to stress here that the requirement for transparency in AI systems is only meaningful when it leads to the end goals of fairness, accountability, or intelligibility.⁸⁹ It is far more effective to embed values of fairness, accountability, and non-discrimination at the time of building AI systems.

The role of the Government

15. ARTICLE 19 believe that the government has a role to play when it comes to AI. In particular, the government should:
 - (i) **Ensure respect for international human rights standards:** A one-size-fits-all approach cannot work in context of the regulation of AI because of the sheer variety of AI systems and capabilities, varying degrees and instances of application, the stakeholders involved, and the nature of decisions being made. However, the minimum requirement for all AI and applications arising from AI should be compliance with international human rights standards.
 - (ii) **Ensure accountability of self regulatory mechanisms:** As ARTICLE 19 has previously stated in its briefing paper⁹⁰ on algorithms and automated

⁸⁵ <https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo>.

⁸⁶ Brent Mittelstadt et al, 'The Ethics of Algorithms: Mapping the Debate', *Big Data & Society*, 3(2).

⁸⁷ Also, see recommendation from the Council of Europe on protection of human rights with regard to search engines here:

https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa87.

⁸⁸ Mike Ananny, 'Toward an Ethics for Algorithms: Convening, Observation, Probability and Timeliness', *Science, Technology and Human Values*, 1-25, 2015.

⁸⁹ Frank Pasquale, *The Black Box Society*, Harvard University Press, 2015.

⁹⁰ Article19’s brief on Algorithms and Automated Decision Making, <http://bit.ly/2f219pd>.

decision making in the context of crime prevention, AI applications are generally used by online intermediaries to block, filter, takedown and remove content. This usually takes place on a self-regulatory basis, though often as a result of government pressure. In practice, this means that wrongful restrictions on access to content are placed beyond judicial scrutiny. At a minimum, the role of the government here is to ensure that individuals have a remedy to challenge decisions based on AI that interfere with their human rights.

- (iii) **Promote a multi-stakeholder approach:** As AI is developed by various actors, and impacts a wide range of actors, policy-making in the area of AI should happen in a multi-stakeholder fashion.

Learning from others

16. ARTICLE 19 notes that the following lessons can be learned from other initiatives:

- (i) **The need for meaningful mechanisms to challenge decisions based on AI:** The European Union's General Data Protection Regulation 2016/679 ('GDPR') provides a right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.⁹¹ The recitals clarify that "in any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision".⁹² While this is a step in the right direction, the scope of what has been dubbed the 'right to an explanation' is limited.⁹³ In particular, it does not apply if the decision: (i) is necessary for entering into, or performance of, a contract between the data subject and a data controller, (ii) is authorised by law; or (iii) based on explicit consent.⁹⁴ It also does not apply if the decision does not produce a legal or similarly significant effect on the data subject.
- (ii) **The importance of aligning policy with technical capabilities:** Policy requirements must be aligned with solutions that are technically feasible, meaningful, and practical. This has been a challenge so far. For example,

⁹¹ Article 22 GDPR.

⁹² See in particular Recital 77.

⁹³ Sandra Wachter et al, "Why a right to explanation of automated decision making does not exist in the General Data Protection Regulation", *International Data Privacy Law* (forthcoming), December 28, 2016.

⁹⁴ Article 22 (2) GDPR.

transparency standards that have traditionally been considered to be a prerequisite to accountability, may not be meaningful or desirable in context of AI systems due to their sheer complexity. Policy requirements must also be practical. For example, the GDPR stresses the importance of meaningful explanations of decisions made by autonomous systems. However, this is in inherent tension with the explanation that machine learning algorithms are developed to offer.⁹⁵

- (iii) **The need to think outside the box:** Considering the new questions that the development of AI raises, for instances in the case of automated vehicles and embodied AI (robots), it is vital to engage in a broad scoping exercise of regulatory efforts around the world. For example, the Committee should take note of the Japanese Robot Strategy,⁹⁶ the forthcoming legal developments in South Korea,⁹⁷ and the calls for a European Agency for robotics.⁹⁸

6 September 2017

⁹⁵Lilian Edwards and Michael Veale, "Slave to the algorithm? Why a right to an explanation is probably not the remedy you are looking for", Duke Law & Technology Review, Forthcoming, 3 July 2017, available at <http://bit.ly/2j10atE>.

⁹⁶"Japan's Robot Strategy was compiled", January 2015, <http://bit.ly/2eF9ZbY>. Also see <http://bit.ly/2vL4frU>.

⁹⁷"Legal preparation for AI era", Business Korea, 17 February 2017, <http://bit.ly/2eEETkx>.

⁹⁸"Robots: Legal Affairs Committee calls for EU-wide Rules", 12 Jan 2017, <http://bit.ly/2x9Wg8m>.

The Association of Medical Research Charities (AMRC) and the Wellcome Trust –
Written evidence (AIC0202)

The Association of Medical Research Charities (AMRC) and the Wellcome Trust – Written evidence (AIC0202)

[Submission to be found under Wellcome Trust](#)

Dr Shahar Avin, Martina Kunz, Andrew Ware, Dr Simon Beard and Dr Seán Ó hÉigeartaigh – Written evidence (AIC0150)

Dr Shahar Avin, Martina Kunz, Andrew Ware, Dr Simon Beard and Dr Seán Ó hÉigeartaigh – Written evidence (AIC0150)

[Submission to be found under Dr Simon Beard](#)

Baker McKenzie – Written evidence (AIC0111)

1. Introduction

- 1.1 Baker McKenzie is an international law firm with 77 offices in 47 countries. We were the first international law firm. We advise many of the world's leading multinational corporations and financial institutions, providing domestic and cross-border legal advice. We employ over 900 people in the UK.
- 1.2 Baker McKenzie welcomes the Select Committee on Artificial Intelligence's inquiry into the implications of advances in artificial intelligence ('AI'). AI is already bringing significant change to multiple industries, including the legal sector and wider professional services. We anticipate that the development of AI will have a major impact on how legal services are delivered in the future. We also believe that law firms can and should make an important contribution to the wider debate on the economic, ethical and social implications of these new technologies.
- 1.3 In this submission we use the term 'AI' to mean the field of computer systems which are able to perform tasks that traditionally require human intelligence without requiring additional programming.
- 1.4 Our submission focuses on three key areas:
 - 1) the importance of education and training;
 - 2) the need for some government intervention, balanced against the risk of over-regulation; and
 - 3) the need for international cooperation.

2. The importance of education and training

- 2.1 AI has the potential to bring significant benefits to society, including greater efficiency in the delivery of legal services. Tools to assist with various tasks currently carried out by lawyers and paralegals, including document review, contract drafting and predicting case outcomes, are already being used by law firms and their use is set to increase⁹⁹. This year Baker McKenzie established a taskforce to investigate the opportunities presented by AI applications, such as machine learning, and to consider longer-term investments in advanced technologies and data management in anticipation of the significant changes these technologies

⁹⁹ <https://www.lawsociety.org.uk/news/documents/future-of-legal-services-pdf/>;
<https://www.ft.com/content/f809870c-26a1-11e7-8691-d5f7e0cd0a16>

will bring to the industry¹⁰⁰. We also advise a number of the leading players in AI technology development.

- 2.2 We anticipate that in the legal sector, as in other industries, there will be a shift over time in job roles, away from tasks which can effectively be automated, such as basic document review and drafting, towards roles which manage and use new technologies to provide creative solutions and nuanced advice. The future we see for the law is machine learning enabled judgement. As an employer, we do not anticipate employing fewer people, but we do anticipate them being more productive and accurate and needing different skills over time.
- 2.3 It is widely accepted that there is a digital skills gap in the UK¹⁰¹. We concur. Education and training will be essential to prepare the workforce to use these emerging technologies effectively. It is therefore important that the government continue to take steps to address the digital skills gap to ensure that the current and future workforce are properly equipped in order to operate in and meet the needs of the new markets.
- 2.4 We appreciate that the UK government has been proactive by promoting STEM subjects and introducing computer science into the school curriculum. However, we believe that there is scope to incorporate digital skills training at all levels of education and across all disciplines, including legal education, and for further training to be provided to persons entering or already in the workforce. In particular, efforts should be made to improve the understanding of, and engagement with, AI applications and wider data handling, manipulation, security, visualisation and analytics skills. Our primary, secondary, tertiary and vocational education needs to have this emergent technology at its core at all levels.

3. Regulation

- 3.1 The development of AI brings with it risks and opportunities. As AI becomes more complex and its use becomes more widespread, we do see a need for some regulation. In particular liability, privacy, control and transparency regimes will need to be considered carefully by legislators and regulators. Rather than reactive regulation (being considered and implemented after something has gone wrong or too far), we recommend proactive, principles-led intervention, based on a sound understanding of the issues and technology, careful consideration and planning. That regulation should be focused on the applications of the technology and deploying the appropriate level of intervention in each case.
- 3.2 Despite the clamour in some sectors of the press for intervention, often based on feared capabilities far beyond the current state of the

¹⁰⁰ <http://www.bakermckenzie.com/en/newsroom/2017/02/innovation/>

¹⁰¹ <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/270/270.pdf>

technology, we recommend staged, considered intervention, with a large element of co-regulation and standard setting as well. We recommend forming cross-interest groups, to include policy makers, consumer and public interest representatives, AI researchers and experts and industry to help further understanding, identify issues, consider guidelines, make recommendations of best practice and consider the impact (intended and unintended) of potential regulation.

3.3 We note that there are already some instances of self-regulation in this area - for example, see the Asilomar Principles¹⁰² and the Partnership on AI's Thematic Pillars¹⁰³. Given the potential for AI to benefit society and the pace at which such technology is evolving, self-regulation should be encouraged. We ourselves have partnered with the World Economic Forum's Centre for the Fourth Industrial Revolution¹⁰⁴ to contribute to research and thinking in this space.

3.4 We also think that that there is a role for government in helping to facilitate ethical (as opposed to legal) frameworks for the development of AI technologies. The right regulatory approach, we have submitted, is staged and considered. But that means that there will be unavoidable gaps where the technology runs ahead of the law. Considered ethical frameworks will help decision makers of all guises make decisions where the law is not yet ready to guide.

4. The need for international cooperation

4.1 Research and advances in the field of AI (and implications of its development) are not limited to the UK. Indeed, several countries in Europe, as well as the U.S.A., Japan, China and Russia are actively working on myriad AI applications. Given the cross-border reach of AI, international cooperation and regulatory harmonisation are crucial.

4.2 Since cross-border legal issues will inevitably arise, they should be anticipated and proactively considered. One topical example is that of the trial of driverless vehicles in the UK¹⁰⁵ which may pave the way for cross-border driverless delivery services in the near future. If driverless vehicles will be potentially crossing borders, neighbouring countries will obviously need to agree how to deal with the adoption of such technology or risk creating inadvertent barriers to trade.

¹⁰² <https://futureoflife.org/ai-principles/>

¹⁰³ <https://www.partnershiponai.org/thematic-pillars/>

¹⁰⁴ <https://www.weforum.org/center-for-the-fourth-industrial-revolution>

¹⁰⁵ <https://www.gov.uk/government/news/over-109-million-of-funding-for-driverless-and-low-carbon-projects>

- 4.3 A report commissioned by Baker McKenzie in 2016, which focused on the role of AI in financial markets, collated responses to a global survey by 424 senior executives from financial institutions and financial technology companies¹⁰⁶. In that report many respondents identified the need for a global approach towards any potential regulation of AI.
- 4.4 We believe it is important for the UK government to lead and engage in dialogue with other nations on the issues related to AI, and believe that international co-operation can facilitate the further development of AI by encouraging shared initiatives, knowledge-sharing and the implementation of best practices.
- 4.5 One way in which international cooperation could be facilitated would be through the set up of an international committee, similar to the International Telecommunication Union¹⁰⁷ or the International Atomic Energy Agency¹⁰⁸ which would help to harmonise standards and approaches as regards AI.

6 September 2017

¹⁰⁶ http://f.datasrvr.com/fr1/516/80536/Baker_McKenzie_Ghosts_in_the_Machine_2016.pdf

¹⁰⁷ <http://www.itu.int/en/about/Pages/default.aspx>

¹⁰⁸ <https://www.iaea.org/about/overview/>

Balderton Capital (UK) LLP – Written evidence (AIC0232)

1. What, in your opinions, are the biggest opportunities and risks in the UK over the coming decade in relation to the development and use of AI?

The growing use of advanced machine learning techniques and artificial intelligence is already creating a fundamental shift in the way software companies operate, and as software becomes the defining asset of many other industries from automotive to finance, applications of AI will fundamentally change how these industries operate over the next decade.

For the UK this is a huge opportunity to apply these new technologies in public services to improve the quality of services and reduce the costs drastically. By supporting skills and building a supportive regulatory environment around the use of these tools, the UK's economy could also benefit hugely by being home to new companies in this space.

However, if we fail to properly regulate these new technologies as they develop, we could see international companies taking advantage of local gaps in our skill base, experience increased unemployment and pressures on social cohesion from the replacement of human task by autonomous services, and even face new threats to our national security.

- i. Are you aware of the Government's recent review? Do you think the Government's AI Review will help with these risks and opportunities? Does it go far enough?

I was aware of the review, predominately through it's promotion by civil servants and ministers on social media. It is an important and timely set of recommendations, and touches on the many ways we can support the development of skills and research around AI in the UK. However the report could do more to cover the regulatory changes that AI may require, such as a touch point in Government for academic, public & private institutions to work with Government on issues in this area as has been done in the past with fintech and other industries.

Overall, it recognises AI as an important area for research and has many promising recommendations to support skill development in this area, but as this comes to disrupt major industries, more direction and engagement from Government will be needed to support it's advantages and curb the threats it poses.

2. What is the general state of AI start-ups in the UK at the moment, what sorts of challenges do you see them facing at the moment, and how do you think these might change over the coming years?

- ii. Is the UK an attractive place to invest in AI? What could be done to improve investment in UK AI businesses?
- iii. Realistically, can UK companies developing AI systems compete with larger companies in the United States and elsewhere?

The UK has a rich heritage as a global leader in the development of AI and is currently amongst the leading ecosystems for such companies in the world. Private funding for such companies in the UK outstrips any other European country, and world leading software companies consider UK-based AI companies some of the best in the world, evidenced in the last few years by multiple partnerships and acquisitions of UK companies and collaborations with our top Universities.

At my fund, Balderton, we deploy millions of pounds each year into software companies based in the UK working on applications of AI as we believe it is one of the best places in the world to do so.

The challenge to us maintaining this position, and rivalling competition that outstrips us in terms of resources and financing in China and the US, is the three-fold.

Firstly investment into academic research in the space must be maintained and broadened to make sure the UK remains a global leader in the field. Up against the many billions of funding targeted directly at the uptake of AI in fields from genomics to robotics by public and private bodies in China, and the huge amount of private capital invested in California into these projects, the UK Government must ensure that our academic institutions have funding available to attract and develop great talent in this area. Secondly access to talent in private companies is essential, and given the huge diversity of nationalities working in this field special visas attached to students studying in this field could be considered to make sure they have the ability to remain in the UK while studying and afterwards when starting companies. And thirdly these companies need to leverage the UK Government's position as a leading digital adopter to give them a competitive edge through access to data and resources in developing new AI use cases.

3. What are the obstacles to AI start-ups scaling up? Is there an investment gap in the UK? If so, how can it be addressed?

The skills required to build competitive AI start-ups today are relatively rare, and as a result the costs for starting a company in this space are higher than other areas of technology. However a drastic drop in the cost of compute and with many parts of the software tooling required being made free by global software companies, this is rapidly changing and by making funds available to train more people in this field, the Government can help increase the supply of labour in this field and thus reduce a major barrier to companies scaling.

The most challenging area of finance in this field is for spin-outs from academic research between launching the company and getting to a first product. AI start-ups have a longer development period due to the complexity of the software involved, and the need for huge amounts of data, resulting in a 'Valley of Death' for start-ups due to the lack of funding before product launch. Post product launch the UK has a healthy number of funds to support these companies, but the committee should look at institutions such as Imperial University for their approach to spin-outs, which could be replicated in other parts of the country, as well as the large private funds investing in Oxford & Cambridge. It would be highly impactful for similar funding and focus to be given to funds attached to other Universities with centres of excellence across the UK.

4. Should more be done to prevent the acquisition of UK AI start-ups by larger foreign corporations? If so, what?

More should be done to make sure that such acquisitions are in the national interest. While we have a takeover panel for strategic assets, it does feel like technology assets are not treated with the same degree of strategic importance to the UK as commodities or even FMCG brands. While, for instance, the Government would rightly oversee a large foreign acquisition of a physical asset based company such as an oil company, the huge value that having large, independent software companies focussing on the development of AI here in the UK seems to be relatively undervalued despite the clear economic and social importance of the UK being home to such companies.

5. Are there any barriers to collaborating with the higher education sector in order to turn AI research into innovative products? What can be done to foster this collaboration?
 - iv. Should taxpayers expect a return on publicly-funded AI research, and if so, what form should that take?
 - v. When public datasets are used to develop commercial AI applications, what benefits should the public expect in return?

Taxpayers should indeed expect a return for investing capital or making data available into research in AI, however this cannot simply be measured in terms of equity value of state or University ownership in these companies. Instead, broader metrics such as job creation, tax receipts, and the multiplier effect of having such companies in our economy should be evaluated on a regular basis to make sure the public are seeing value for money over a suitable time period.

We are generally not convinced that direct state ownership in private companies is beneficial for small technology companies due to the conflict of picking winners in such an early stage of a new industry. However state support in the funding of research, crafting of regulation and promotion of a supportive ecosystem is

invaluable, so the more the Government can do, such as making public datasets available, the better.

6. Do investors have a duty to ensure that AI is developed in an ethical and responsible way? If so, how should they follow through on this?
- vi. What are your views on the use of regulations or voluntary measures to ensure AI is developed and used in an ethical ways?

Board members do have a duty to the company, employees and society in which the company is based to ensure ethical behaviour of any company where it is possible. This is no different in AI, except to recognise that the techniques used in the development of many forms of AI are extremely difficult to dissect vs oversight of employee behaviour or more standard codebases and guidance could be offered to smaller companies in this space, as well as to investors, here.

However, as 'AI' is not itself a defined sector but rather a tool applied to existing services, the ethical use of such tools would be best regulated by the specific industry rather than having an independent ethics committee to rule on issues as broad as the application of such technology in financial trading to medical assessments. By pushing for a cross Government body, similar to the Open Data project run by Government, different regulators, investors and companies could share views and provide input into regulations that are relevant to their industry, but learn from other areas of best practice across sectors.

7. Should individuals expect to be able to retain ownership of their personal data and still benefit from developments in AI, or are these two aims incompatible?

Individuals should be made fully aware, and asked permission for the use of their data for third party use where they can in any way be individually identifiable. Adoption of GDPR rules will be a good start with this, and companies such as Patients Know Best provide good examples of how permissions can be acquired through opt ins procedures.

However, to support the development of new AI companies, the public sector should make any non-personally identifiable information available to third parties and the recommendation of data trusts in the Government's AI review is a powerful start.

James Wise, Partner, Balderton Capital (UK) LLP

20 November 2017

Professor Andrew Basden – Written evidence (AIC0195)

6 September 2017

I am pleased to contribute this is the submission to The House of Lords Select Committee on Artificial Intelligence, where I am grateful for the help provided by colleagues in the UK Christian Academic Network (C-A-N-), who have a growing interest in facilitating contribution to the discipline as detailed in the preamble below . I am Professor of Human Factors and Philosophy in Information Systems at the University of Salford, Business School. Evidence provided in this contribution originates from over 35 years of experience relating to the field, in both industry and academia, where I have authored/co-authored over 70 publications related to issues mentioned below, and have completed a soon-to-be-published book: *Foundations of Information Systems: Research and Practice*. The next page constitutes an introductory page to a framework I have detailed in this document, in which reference is made to questions posed in the *Call for Evidence*.

PREAMBLE FROM THE CHRISTIAN ACADEMIC NETWORK

The Christian Academic Network (C-A-N-) was formed in 2001, with the vision to proclaim the gospel of Jesus Christ through the academic community of the UK. In fulfilling this vision, we have two major aims: A) to encourage the integration of Christian faith into academic life and B) to support and equip all university and college staff as witnesses for Christ in their workplace. We recognise that the contributions of Prof Andrew Basden may or may not reflect the views of others within the C-A-N- membership. As the House of Lords Select Committee on Artificial Intelligence is in formation at this time, this is very timely for our organisation where we are actively seeking to support Christian academics with interest in Artificial Intelligence, which we hope will make useful future contributions to the committee, which to our knowledge are currently not covered by existing discipline specific groups in the UK.

We seek to encourage the integration of Christian faith into academic life in order to support and equip University and College staff and those academics working in research as witnesses for Christ in their workplace. A practical way in which this is facilitated is through our work to support Christian academics in finding a way to critically affirm and enrich the thinking in their fields, as opposed to taking an antagonistic position.

In light of this approach, what we would support academics to offer, is not a critical view of the discipline of Artificial Intelligence, but a framework within which all from different backgrounds can critically review and embrace all topics currently of interest to the committee. We recognise the contribution of Prof Basden contained in this response to fit within this context. The help others on

the C-A-N- executive in articulating the presentation of the contribution is also acknowledged.

INTRODUCTION

1. To consider "the economic, ethical and social implications of advances in artificial intelligence" requires a broad view, in which all issues are able to be seen in relation to all others. Any aspect of AI that is neglected might be the one that jeopardises the whole, especially in the longer run.
2. So this submission does not address specific issues, which are many, varied and the discourses around which are often isolated from each other. Instead, it offers an **integrating conceptual framework**, which addresses six broad areas of discourse around AI (explained below).
3. Key. Throughout this document, "Fn", "Fn.m" etc. refer to items in the framework outlined below. "Qn" refers to a question posed in the Call for Evidence.
4. Within this framework, most issues of interest to the Committee, and the questions the Committee posed in their Call for Evidence, may be situated and understood in relation to the others. This offers a holistic conceptual toolkit with which the Committee might systematically (a) clarify the complexities of specific issues in contributions they receive, (b) understand their background concerns, (c) discern relationships among them, (d) understand more clearly the range of interest, and (e) detect issues that have not been covered.
5. The framework seeks the peace and wisdom embodied in the Hebrew word *shalom*, of which prosperity, justice, ethics, human flourishing and environmental responsibility are all integral parts working in harmony.
6. **Definition of AI.** From my experience, I see AI as that subset of computer science and information systems which acts according to expert knowledge, which is stored in knowledge bases and algorithms for reasoning or learning. The definition of AI cannot be without reference to its context of use and beliefs. Its roles range from automated decision-making and action, through advice-giving to enhancing and stimulating human thinking [Note 1].
7. There are **two main kinds of AI system**: (a) knowledge-based or expert systems, in which knowledge that has been acquired by human knowledge elicitation methods, is represented explicitly in the computer, (b) machine-learning systems, in which knowledge is built up from automated analysis of many training examples, this learning being directed by human-selected parameters, which might be complex. 'Black-boxing' (Q9) occurs especially with machine-learning approaches, whereas explicitly-represented knowledge can be transparent.

EXPLANATION OF THE FRAMEWORK

8. The integrating framework incorporates six main areas of concern in relation to AI, as depicted in Figure 1 [Note 2].

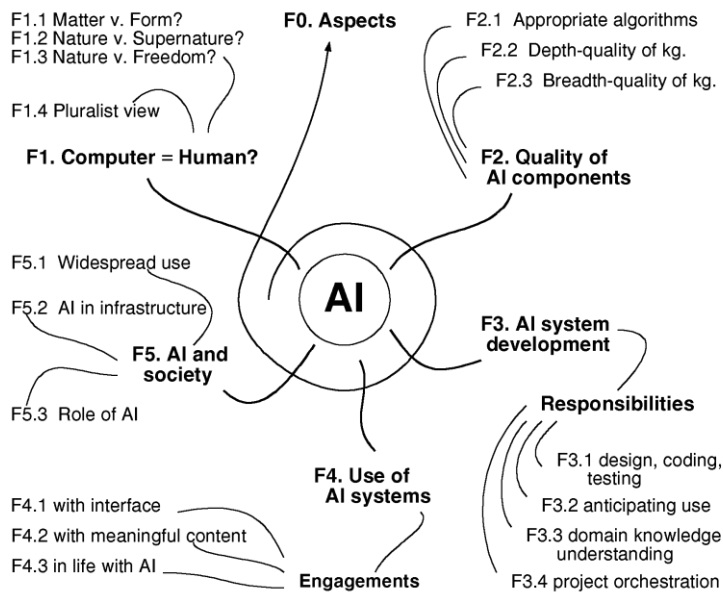


Figure 1. Framework for Understanding Artificial Intelligence

9. The framework may be used to prepare the public (Q3, Q5): it defines areas of concern in each of which public discussion can be encouraged.

F0. Which Aspects are Important?

10. The ethical, social and economic implications of AI involve multiple aspects, some of which are frequently overlooked. It is therefore important to be aware of possible aspects of the world, of which humanity, AI, society and the planet are part. Maslow's hierarchy of needs offers one suite of aspects, but the following is more comprehensive, is philosophically sound and has been employed in the information systems field [Note 3].

- Quantitative: quantity, amount
- Spatial: continuous extension, space
- Kinematic: movement; flowing movement
- Physical: energy + mass
- Biotic: vitality, health v. disease
- Psychic: sense, feeling, emotion v. insensitivity
- Analytical: distinguishing, conceptual clarity v. confusion
- Formative: history, culture, technology, goals, achievements, shaping, industry v. laziness
- Lingual: symbolic communication, understanding v. deceit
- Social: togetherness, organisation, friendship v. enmity
- Economic: frugal management of resources v. waste
- Aesthetic: harmony, interest, fun v. chaos, boredom
- Juridical: what is due; appropriateness, right, responsibility, justice v. injustice, inappropriateness

- Attitudinal / Ethical: attitude of self-giving love, generosity v. selfishness, self-protection
 - Pistic: faith, beliefs, aspiration, commitment, religion, ultimate meaningfulness: courage v. cowardice
11. Each aspect is irreducible to others, yet depends intrinsically on the others. Most define a different normativity (good, evil). Notice that the ethical aspect is to do with attitude, not right-and-wrong, which is juridical; privacy is a juridical issue, societal diversity is partly attitudinal.
 12. For the Committee, this suite of aspects can offer a useful conceptual tool for clarifying issues around the being, activity and normativity of AI. Example: Reasons for current **excitement in AI** (Q2) may be understood via the aspects (e.g. technical, economic, inspirational). Most issues are multi-aspectual, e.g. **transparency** or black-boxing (Q9) is meaningful in both ethical (self-protection) and formative (technical: kind of AI) aspects.
 13. The **role of Government** (Q10) is primarily centred on the juridical aspect (legislation). However, Government also has a *de-facto* role of leadership in society, alongside the media, academia, religious institutions, etc. Leadership in society concerns the ethical aspect (pervading attitude) and pistic aspect (setting aspirations etc.). So I draw the Committee's attention especially to the juridical, ethical and pistic aspects in all areas of concern.

F1. The philosophical question: "Computer = Human?"

14. Our view of AI depends on how we approach the philosophical question of whether computers can be like human beings, now or in the future. Whereas philosophers debate this, most AI discourses presuppose one basis tacitly, which can also be its inspiration (Q2).
15. Four possible bases for addressing that question may be detected, in which the AI question takes on a different form.
 - F1.1. Matter-form dualism. "Computer is matter. Human mind is form. Can form be reduced to / emerge from matter, in either substance or behaviour?" Example: Several attempts to address the question found in Boden [1990].
 - F1.2. Nature-supernature dualism. "Computers are natural. Humans are 'supernatural'. Can the supernatural be reduced to the natural?" Examples: Scholasticism ('supernatural' = 'divine spark'), Searle [1990] (supernatural = biotic rather than physical causality).
 - F1.3. Nature-freedom dualism. "Computers are determined. Humans are free. Is freedom illusory?" Example: Newell [1982]. Alternative question: "Humans are subject. Can computers become subject rather than object?"
 - F1.4. Pluralistic aspects. "Computers and humans both function in the same aspects of reality (F0). Humans function as subject in all aspects. In which aspects may computers function as subject?" Example answer:

computers function as subject up to physical aspect, as subjects-by-proxy in psychic to lingual aspects, and as object in social to pistic aspects [Basden 2017] (See F0).

16. The three dualistic bases tend towards conflict in debate, the fourth, towards resolving the conflict, in that it allows multiple answers to co-exist. F1.4 treats ethicality (Q8) as innate, rather than bolt-on.
17. Discerning which basis is presupposed in society, has implications for the kinds of principles on which Governments enact policy and legislation, because it influences our view of what it means to be human and our relationship with the rest of Creation.

F2. Quality of Components from which AI applications are constructed

18. AI systems work by algorithms acting on a knowledge base. These exhibit three kinds of quality, each with attendant dangers for AI systems behaviour:

- F2.1. Appropriateness of algorithms, protocols, etc. being employed. Danger of errors or of short-cut algorithms.
- F2.2. Depth-quality of semantic content of knowledge bases. This is completeness and accuracy of knowledge of laws within single aspects. Example: Deep Mind defeats Go champion because of its deep knowledge in spatial, formative and a few other aspects. Danger of exceptional conditions not being included because they are rare and/or tacitly known. Example: Year-2000 bug.
- F2.3. Breadth-quality of the represented knowledge or parameters that control them. This refers to whether all relevant aspects have been adequately represented among the parameters. Danger of omission or over-emphasis of aspects (deliberate or inadvertant), yielding biased, distorted results. Example: economic parameters included, ethical parameters omitted.

19. These apply to both kinds of AI, whether humanly-represented or machine-learned. In the latter, depth-quality is determined by number of training examples, and breadth-quality by the parameters that are set for learning.

20. Re. Q1: AI can take the automation role more successfully when its knowledge is narrow, especially when in one aspect (Example: Deep Blue, controllers of self-driving cars). In multi-aspectual situations, AI is unlikely to ever take an automation role, but might be successful in the stimulation of human thinking [Note 1]. Example: This implies that in self-driving cars, where the automation role is central, the depth-quality knowledge of the spatial, kinematic and physical aspects is crucial to general safe driving. However, if the AI must decide whether to crash into child or adult, then complex, multi-aspectual knowledge is required (breadth-quality), which is unlikely ever to be possible.

21. The role of Government here (Q10) is less likely to be legislative, and more likely to be as leader in society, alongside other sectors, to shape attitudes and aspirations. However, one possibility for legislation is that all AI be accompanied by a substantial publicly-available, written explanation of the knowledge contained, including the parameters used for learning at all times, thus increasing transparency.

F3. Development and Supply of AI Systems from these Components

22. AI system development and supply involves four intertwined activities, each with its distinct responsibility:

- F3.1. Creating the AI system; responsibility for careful design, coding, testing;
- F3.2. User requirements analysis; responsibility for engaging all relevant stakeholders, and for properly anticipating situations of use, and their possible benefits and harm in use;
- F3.3. Domain knowledge (or parameter) elicitation; responsibility for eliciting complete and accurate knowledge or parameters, including all relevant aspects and their laws;
- F3.4. Project management; responsibility for orchestration of all considerations, in all aspects. Example: arrogant attitude in development team can lead to dismissal of valid concerns.

23. All four responsibilities involve multiple aspects (F0), such as technical, lingual, social, jural, ethical. Recognition of all the distinct aspects of each responsibility is crucial to successful and non-harmful application of AI. This offers insight into complex 'ethical' issues (Q8), by analysing which aspects of which responsibilities are involved. It is appropriate for the Government to legislate in regard to all four responsibilities, especially around duty of care.

F4. The use of AI systems

24. People engage with the AI systems in three main ways, each with a main norm.

- F5.1. Engagement with the interface and technology (mouse and screen, virtual reality headsets, direct nerve connections, etc.) Norm: ease of use; should usually be 'immersive'. In embedded systems, this engagement is at most rudimentary.
- F4.2. Engagement with meaningful content; especially important in games and virtual worlds. Norm: veracity; the AI of the virtual world must be 'realistic'. Both addiction and information satisfaction arise here.
- F4.3. Engagement in life with AI, which brings benefit or harm in life and work or users and others, which might be indirect, e.g. effect of game-addiction on family. Norm: usefulness or beneficiality.

25. All three involve all aspects in principle (F0), both their benefit and harm. Impact on everyday life (Q3) and gainers and losers in AI (Q4) may both be analysed systematically via aspects of F4.3 (benefit and harm). That some aspects might be positive and others, negative, offers nuanced analysis.
26. Aspectual analysis of each engagement can help the Committee clarify and tease out multiple impacts, and, when used with imagination, reveal issues often overlooked. Example: Can a person have a relationship with of self-giving love (as opposed to entertainment) with a care robot as they do with a dog? Legislation for F4.3 may be effective, but unlikely for F4.1, F4.2.

F5. Relationship Between AI and Society

27. There are several circular, interacting relationships between AI and society.
- F5.1. Impact of widespread use of AI systems (F4.3 amplified). Impact on behaviour (e.g. AI in gambling), on attitudes (e.g. AI in social media), on aspirations (e.g. in online marketing), all contributing to public good (Q7). AI is relevant to climate change emissions, as people are led to aspire to things with more, or less, carbon footprint.
 - F5.2. AI as (part of) infrastructure. Each aspect provides a different infrastructure in which AI plays a part: technical infrastructure (e.g. AI in search engines) is of formative aspect, economic infrastructure (e.g. AI in share dealing), etc. Two infrastructures are often overlooked: (a) ethical infrastructure of attitudes that pervade society (e.g. "Winner takes all", Q7), (b) pistic infrastructure of prevailing beliefs, aspirations and assumptions; both are shaped by AI-controlled newsfeed etc. [Note 4].
 - F5.3. The role of AI in society, and progress in which AI has a part. Progress may be seen as the opening up of the potential of the aspects and, though technology has its own norm, this should be guided by the norms of all other aspects [Note 5]. There should never be "AI for AI's sake".
28. Danger of whole constellations of issues being suppressed and never discussed. Therefore in an exercise like this consultation, it is important to ensure that issues in all aspects are given their due.
29. "Ethical, social and economic implications" of AI (Q8), in their societal manifestation all relate to the (post-)social aspects, as widespread impact (F5.1), these aspects of infrastructure (F5.2), and of progress (F5.3). Key sectors (Q6) relates to F5.2, F5.3. Note: many so-called "ethical" issues (Q8) are actually juridical, to do with rights, justice, etc., and should be distinguished from issues of attitude.
30. The role of Government (Q10) is juridical for legislative infrastructure, but is as society-leader in ethical and pistic infrastructures, and for these legislation is less effective.

CONCLUDING REMARKS

31. This submission suggests a framework by which disparate discourses around AI and the issues they find meaningful may be situated within an integrated understanding of AI. This framework addresses six main areas of concern; others may be added if desired.
32. AI in society cannot be properly understood if any area is ignored. Each area links with others. Awareness of aspects (F0) pervades all other areas. F1 determines what society expects of AI (F5), which applications emerge (F3, F4), and which aspects are deemed meaningful or neglected (F0). AI components (F2) supplies F3 and also enables and constrains F4 via affordance and appropriateness. F3 delivers F4 which, if widespread, results in F5. F3 also influences which components (F2) are designed. F4 is determined by F1, constrained by F2, a result of F3, and can generate F5.1. F5.1 is amplified F4.3; F5.2 constrains F3, F4. F5.3 is affected by F1.
33. By analysing how individual issues relate to each area, together with which aspects are important, the issues can be (a) clarified, (b) linked with other issues, (c) seen in context, (d) understood as innately normative (ethical) in nature. It might offer a sound basis for suggesting issues that are currently overlooked and might appear in future.
34. The Appendix collects together suggestions above for how the framework may be used to address each of the questions posed in the Call for Evidence

NOTES

- Note 1. Eight roles of expert systems, which arose in practice, were discussed by Basden [1983]. The automation role is exemplified by the controllers of self-driving cars and by Deep Mind. The advice-giving role might be found in Google Search. The role of stimulating human thinking and discussion was found in the ELSIE expert system, which helped quantity surveyors think through the advice they gave to clients. The *Assistum* software (www.assistum.com) was developed from an expert system in ICI plc that encouraged users to question it.
- Note 2. The framework is adapted from those discussed in Basden [2017], which are based on the Reformational Philosophy of Dooyeweerd [1955]. It discusses much of the practical relevance of a general version of the framework for understanding ICT and, in its final chapter, makes over 100 suggestions for improving practice and for research projects.
- Note 3. This suite of aspects is from Dooyeweerd [1955]. See Basden [2017] for discussion of its strengths and applications.
- Note 4. A circular relationship in which agency reconstitutes structure, which influences agency (Giddens' structuration), but in multiple aspects.
- Note 5. Schuurman E. 1980. *Technology and the future: A philosophical challenge*. Toronto, Ontario, Canada: Wedge.

REFERENCES

- Basden A. 1983. "On the application of expert systems", *International Journal of Man-Machine Studies*, **19**, 461-477.
- Basden A. 2017. *Foundations of Information Systems: Research and Practice*. Routledge.
- Boden, MA. 1990. Escaping from the Chinese room. In M. A. Boden (Ed.), *The philosophy of artificial intelligence* (pp. 89-104). Oxford, England: Oxford University Press.
- Dooyeweerd H. 1955. *A new critique of theoretical thought (Vols. I-IV)*. Jordan Station, Ontario, Canada: Paideia Press. (Original work published 1953-1958)
- Newell A. 1982. The knowledge level. *Artificial Intelligence*, **18**, 87-127.
- Searle JR. 1990. Minds, brains and programs. In M. A. Boden (Ed.), *The philosophy of artificial intelligence* (pp. 67-88). Oxford, England: Oxford University Press. (Original work published 1980 in *Behavioural and Brain Sciences*, **3**, 417-424)

APPENDIX

This appendix outlines how the framework might be employed to address the questions that are included in the Call for Evidence.

Question 1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- Discourses around AI are fragmented, with those related to ethics of AI (F4, F5) disconnected from technical (F2) or philosophical (F1) discourses. This framework can bring them together.
- Instead of trying to predict how AI is likely to advance, we suggest trying to steer its advance, especially in its use.
- The pistic and ethical aspects are particularly important in steering the advance of AI, and their impact is long-term? Earlier aspects have shorter-term impacts.

Question 2. Is the current level of excitement which surrounds artificial intelligence warranted?

- AI should be seen as emerging from the potential with which the Creator has invested the Creation, but also the responsibility that He laid on human beings to "tend and care for" it (Genesis 2:15).
- This care should relate to every aspect (F0).

Question 3. How can the general public best be prepared for more widespread use of artificial intelligence?

Question 5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

- The framework may be used to prepare the public: it defines areas of concern in each of which public discussion can be encouraged.
- Impact on everyday life is the concern of F4.3, on jobs is a concern of F5.1. Skills is a concern of F2, F3. Social policy changes are a concern of F4, F5.
- Consideration of these questions requires not just public discussion (lingual aspect) nor facilitation (economic aspect), but also attention to the effect of aspirations and attitude (ethical, pistic aspects). F0.

Question 4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated? , about gainers and losers.

- Who is gaining most and least from AI depends on from which aspect we look (F0), and relates to F4.3. If this is amplified by widespread use (F5.1), this can become a major concern.
- Disparities are problematic because of the juridical aspect, and tend to occur because of the attitudes and beliefs that pervade society (ethical, pistic aspects). F0.

Question 6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

- Likewise, key sector identification may be facilitated by reference to F4.3, considering each and every aspect in turn (F0).
- The role of AI as serving other aspects rather than itself (F5.3) becomes important.

Question 7. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- The pistic aspect is the most fundamental here, affecting humanity's functioning in all areas. Christians believe that without recognition of responsibility to the Creator, the 'public good' is ill-defined.
- "Winner-takes-all" is gross dysfunction in the ethical aspect of attitude (F5.2), which undermines society. Ethical-aspect self-giving is what society's leaders should be modelling and advocating.
- Contributing to "a well-functioning economy": Beware of seeing 'the economy' as the only important aspect of life.

Question 8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

- Every part of the framework offers different understanding of 'ethical' implications of AI.
- In considering 'ethical' issues, be aware of the distinction between juridical aspect (right, wrong) and ethical aspect (attitude). Privacy, consent, safety, and impact on democracy are issues of the juridical aspect, while diversity is an issue of attitude. Moreover, take account of the distinct normativity of each aspect (F0).
- F1, the philosophical view of the basis of the AI question, influences our view of ethics. F1.4 makes ethicality innate rather than bolt-on.
- Development of AI: Ethical issues occur in the four responsibilities of developers (F3). F2 emphasises the responsibility for good quality knowledge and other components.
- Use of AI: F4.3 concerns individual and organisational use of AI, while F5.1 concerns widespread impact of this use in society. F5.2, concerning the 'attitude infrastructure', influences how both of these are seen, and is set by leaders in society (including Government).

Question 9. In what situations is a relative lack of transparency in artificial intelligence systems (socalled "black boxing") acceptable? When should it not be permissible?

- 'Black-boxing' occurs especially with machine-learning approaches, because the built-up knowledge is not explicit. This contrasts with representational AI, where knowledge is explicit and transparent. F2.
- Advantages and disadvantages of black-boxing are different in each aspect (F0). Black-boxing increases technical efficiency (economic benefit), reduces understanding (lingual harm), implies reduced generosity towards users (ethical harm).
- Suggestion: Could policy be enacted that all machine-learning AI be accompanied by a substantial written explanation of the knowledge contained, combined with transparency over the parameters used for learning? Especially important where safety is important (e.g. self-driving cars).

Question 10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- The role of Government is primarily qualified by the juridical aspect. It is not the role of legislation to attempt to set morals or beliefs in place. Any attempt of legislation to do this is usually ineffective and counter-productive.
- However, the role of those human beings involved in Government is wider, including societal leadership in ethical and faith aspects (especially in their manifestation as attitude and aspiration).

Question 11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

- To answer this, consider each area of concern in turn (F1 - F5) and each aspect thereof (F0).

More detailed outworking of the Framework may be found in chapters 5 to 9 of Basden [2017].

21 September 2017

BBC – Written evidence (AIC0204)

1. The BBC welcomes the appointment by the House of Lords of a Select Committee on Artificial Intelligence (AI) and the opportunity to submit written evidence to its inquiry. We are responding as a significant contributor to the UK's success in engineering and digital products, with a Royal Charter mission to promote technology innovation and skills. In this response, we have focussed particularly on the themes of public understanding and safeguarding, and the ethical implications of the development of AI.
2. AI systems - particularly machine learning systems - that learn, and which can act autonomously, are developing rapidly and signify a paradigm shift in the relationship between society and technology. In this response, we acknowledge that AI systems have the potential to deliver significant benefits to all members of society and promise to transform the creative capabilities and productive capacity of organisations. But we also set out a number of significant challenges to society posed by AI in terms of the effects of increasingly pervasive use and control of AI and personal data, and the impact of AI on employment.
3. We go on to discuss the importance of creating the right ethical, policy, public engagement and regulatory environment for the development of responsible AI and the role that the BBC, guided by its public service principles, can play. We conclude that AI can have a significant and positive impact on society, and that the UK can be a leader in this field provided we are guided by public interest. We also make the case for the BBC's critical contribution, as part of a mixed AI ecosystem, to the development of beneficial AI both technically, through the development of AI services, and editorially, by encouraging informed and balanced debate.

Artificial Intelligence and Society

4. Media reporting on AI has increased significantly over the last year partly in response to profound developments in the field. Access to data, cloud computing, the best talent, and large amounts of capital have allowed a few businesses using AI to develop in ways that were just not possible three to five years ago, and we expect more businesses to adopt AI to power their growth. Recent analysis by the Royal Society¹⁰⁹, and

¹⁰⁹ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

McKinsey¹¹⁰ has highlighted almost all industries stand to be impacted by AI. There is good evidence of potentially large societal and economic benefits from AI in terms of, for example, better healthcare, greater access to information and more efficient workplaces¹¹¹. While the potential for good with AI is significant, it is also important to highlight possible challenges which require open and robust debate. We have summarised them under four headings below:

- 4.1. **The pervasive role of AI in mediating our digital lives:** In the near future, it could become impossible to opt-out from AI. The deployment of well-intentioned AI systems of support – designed to aid and augment society – risk becoming systems of control¹¹² if due care is not taken to assess when and how these systems are used to determine life changing outcomes for individuals¹¹³. We are also seeing some evidence that individuals are becoming locked into digital platforms/ecosystems¹¹⁴¹¹⁵¹¹⁶ which increases the chance of monopolies occurring, and there is a danger the situation will become more acute as AI starts to mediate our public and private lives. We can address this by opening up choice for the consumer and building in means for transparency and review from the start.
- 4.2. **The use of AI to influence behaviour:** The predictive and analytic capability of AI (for example to complete a web search, automatically respond to messages, draw inference from a wide range of data, or to offer personalised recommendations) is of great utility. However, AI systems that shape and direct the public’s attention risk straying into social engineering¹¹⁷. AI will come to control the information we see and the choices offered to us, and there is real worry over the role AI (and the organisations controlling AI services) will play in shaping the norms and values of society.

¹¹⁰ <http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-artificial-intelligence-can-deliver-real-value-to-companies>

¹¹¹ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

¹¹² <http://www.bbc.co.uk/programmes/b091wb34>

¹¹³ <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

¹¹⁴ <http://www.bbc.co.uk/programmes/b08fgvln>

¹¹⁵ <https://www.nytimes.com/2017/03/21/magazine/platform-companies-are-becoming-more-powerful-but-what-exactly-do-they-want.html>

¹¹⁶ <https://www.theguardian.com/technology/2015/jun/07/facebook-uber-amazon-platform-economy>

¹¹⁷ <http://www.bbc.co.uk/programmes/p056p7wb>

- 4.3. **The displacement of the workforce:** The impact of AI on jobs will be both positive and negative. On the positive side, we expect AI to provide tools for analysis and decision making that improve productivity and allow businesses – including the BBC - to achieve things they would otherwise have been unable to do. But on the negative side, recent analysis by PwC¹¹⁸ suggests that 30% of UK jobs are at risk of automation by the 2030s. The scale and pace of disruption across a wide range of industries at the same time is likely to cause significant societal impact, in ways that previous transformations of one particular industry have not. While AI is expected to impact both white and blue collar jobs, we are concerned that the most vulnerable in society will suffer the most disruption to their employment due to AI. A US report published in Dec 2016¹¹⁹ highlighted the risk of inequality increasing with AI. Any discussion on the development and introduction of AI in society must address how this disruption and inequality is mitigated.
- 4.4. **The use of individual's data:** The collection, preparation and use of data in AI will be central to the future effectiveness and fairness of the AI ecosystem. Data privacy, security of data, and how bias (explicit or implicit) in data¹²⁰ is dealt with, will define the extent to which AI acts as a positive force in society. Today, large amounts of personal data are controlled by a few organisations, resulting in disproportionate influence and control. This threatens to limit competition and external innovation, and deny citizens control over what is fundamentally theirs. Those who are least well informed in society may find themselves disempowered by the automated decision-making of machine learning. For example, poorer, older and less well-educated members of society may be unable to influence how their personal data is managed, or understand their rights in an AI powered society. For AI to be a success both in market and societal terms, it is essential that citizens understand what information has been collected about them, how it has been used to train AI systems, and how those systems will then make decisions that impact them.

Creating the right environment for AI

5. The challenges listed above demonstrate how AI developments left to emerge organically in an entirely unregulated free market could develop major problems for citizens, and leave them vulnerable to the power of large corporations. It is clear that the complexity and enormity of these challenges

¹¹⁸ <https://www.pwc.co.uk/economic-services/ukey/pwcukey-section-4-automation-march-2017-v2.pdf>

¹¹⁹ <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>

¹²⁰ <https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals>

cannot be resolved by technology innovation alone. Just as with any market we must consider the appropriate approach to regulation, data access, employment, and the engagement of society at large. It is essential we create the right environment for AI, to ensure it delivers the most value to citizens, and provides a stable commercial environment in which innovation can flourish.

6. Others have already begun this debate. The Royal Society and the British Academy for example has suggested the formation of a stewardship body and four principles for data governance: protect individual and collective rights and interests; ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively; seek out good practices and learn from success and failure; and enhance existing democratic governance¹²¹. We agree with this and suggest three further key tenets:

- 6.1. First, **the regulations in affected industries need to be fit for AI**, and able to evolve as it develops. The rapid development of AI requires lawmakers and regulators to keep any AI framework under review and up-to-date, and to consider how any regulatory framework can be applied consistently globally.

- 6.2. Second, **individuals must have ownership and control over the data collected** from them -directly or indirectly. Individual's data should be portable – able to be easily moved from provider to provider - and should remain the property of the individual and not the collecting organisation. New legislation on data portability and the work of the Information Commissioner's Office (ICO)¹²² are therefore welcome. We feel it is important for organisations to meet the spirit as well as the letter of the law, and practical data portability, privacy and protection should be designed into AI services from the outset. It is also important that we innovate when it comes to managing data, and learn from the experience of others - approaches to data management (for example Differential Privacy by Apple Inc¹²³, the Second Payment Services Directive¹²⁴, and the FCA Sandbox¹²⁵) offer valuable lessons on how to open-up access to data, and manage the challenges in a responsible, equitable way.

¹²¹ <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf>

¹²² <https://ico.org.uk/for-organisations/guide-to-data-protection/big-data/> ; <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

¹²³ <https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/>

¹²⁴ <https://www.pwc.co.uk/industries/banking-capital-markets/insights/psd2-a-game-changing-regulation.html>

¹²⁵ <https://www.fca.org.uk/firms/regulatory-sandbox>

- 6.3. Third, and finally, if individuals are to build confidence in the role of AI in their lives and society more broadly, then **they must understand how AI affects them**. When personal data is used by AI to make decisions that will impact an individual, then it should be the right of an individual to understand on what basis decisions about them are made. While we accept that the mathematical models underpinning AI can by their very nature be a 'black box', it is in our view no longer acceptable for algorithms that control decisions about an individual to be entirely hidden behind commercial confidentiality or technical convenience. Recent academic research presented at the Leverhulme Centre for the Future of Intelligence¹²⁶ has shown it is possible to expose the characteristics of the decision-making without needing to peer into the black-box itself – that is, it is possible to offer algorithmic transparency without having to compromise intellectual property.

A public service approach to AI, and the role of the BBC

7. Investment and development of AI by the commercial sector is welcome - businesses across the world are making significant and, at times, high-risk investments in AI which have the potential to drive significant consumer benefits. But, for the full benefits of AI to be felt by society, it's important that AI services are shaped by forces wider than those based in Silicon Valley or Beijing¹²⁷. It is no coincidence that many of the leading thinkers and practitioners of AI are of UK origin. They may now be powering development for the global corporate giants, but it was investment in the UK education sector that nurtured the UK's world class talent in the first place.
8. Such a significant part of our future society requires an open debate, engaging not just engineers but social and political scientists; economists and anthropologists; journalists, and creatives; and - most critically - the general public. The BBC is well positioned to spark this debate and to set standards for technical development of AI services. We bring not only breadth and depth in engineering, including a significant digital research and development capability, but also editorial and creative expertise, and a history of making audience voices heard.

¹²⁶ <http://lcfi.ac.uk/about/>

¹²⁷ <https://www.economist.com/news/business/21725018-its-deep-pool-data-may-let-it-lead-artificial-intelligence-china-may-match-or-beat-america>

9. The BBC’s mandate to inform, educate and entertain is underpinned by a set of guiding principles which we believe are equally relevant to the BBC’s development of its own AI services and to its approach to the editorial coverage of AI. We would also suggest that they can be adapted by other organisations for the wider development of AI beyond public service. The principles are:

- 9.1. **Independence** – it will be increasingly vital that people can find information and recommendations they can trust and that they can be sure are free from commercial or political agenda
- 9.2. **Impartiality** – we should promote services that are built to minimise the bias (implicit or explicit) that can arise from training machine learning on data that reflects existing prejudice or has been developed by designers that fail to reflect the diversity of society¹²⁸
- 9.3. **Accountability** – With decision-making buried deep within the AI process, it is even more important to ensure that providers of AI services remain accountable to their users in a meaningful way
- 9.4. **Universality** – We must avoid an AI future that is limited only to the wealthy or well educated few, or even one in which AI services are limited to a small number of companies who have exclusive access to the data

10. With these principles guiding our way, we want to highlight three main ways in which the BBC can play a public service role in the development of the right environment for AI:

10.1 **Informing the Debate** – The BBC reaches around 95 percent of the UK population every week, and as many as 370 million people across the globe. BBC News is the most trusted source of news in the country¹²⁹. As the national broadcaster with a global reach, our responsibility with any issue that is going to have such profound and far-reaching implications for society is to help make sure there is a truly informed debate. We have already demonstrated that we take that responsibility seriously through our content on TV, radio and online. In recent months, the BBC has covered stories ranging from the role AI can play in improving cancer diagnosis¹³⁰, to the open letter by 100 experts to the UN in relation to AI and lethal autonomous weapons¹³¹, and the success of Google’s Alpha-Go AI at the strategy game

¹²⁸ <https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals> ; <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

¹²⁹ https://www.ofcom.org.uk/data/assets/pdf_file/0016/103570/news-consumption-uk-2016.pdf

¹³⁰ <http://www.bbc.co.uk/news/health-38717928>

¹³¹ <http://www.bbc.co.uk/news/technology-40995835>

Go¹³². Perhaps even more important is attention the BBC has given to developments that are quietly shaping society today, enabled by AI but largely hidden from public view. For example, the BBC has reported on the increase in gig economy platform services¹³³, dynamic pricing in retail¹³⁴, personalisation of social media¹³⁵, impact on the law¹³⁶, transport¹³⁷ and health care¹³⁸. The continued independence of our editorial output will ensure UK and global audiences can trust the BBC to support impartial, balanced debate on this subject.

10.2 Bringing Partners Together – The BBC will play an active role in improving the level of public understanding about AI and its effects on society by using our unique convening power to bring together leaders from public service institutions, academia, and the commercial sector around the biggest issues, and by sharing our combined knowledge and acting as a trusted impartial bridge between communities with different perspectives. We have started to do this both in public forums such as conferences and symposia, and in specialist partnerships, such as the Data Science Research Partnership – where we are working alongside eight universities to help shape the future of research into data science for public good.

10.3 Responsible Technical Development – The BBC can engage the public in the development of AI and make them informed consumers, through the BBC’s use of AI technology in the services we provide. While so far the BBC’s application of AI is relatively utilitarian in nature we anticipate that in the next few years machine learning will support and underpin every aspect of our audience offer. The BBC must use its significant digital and engineering R&D capacity to be a pioneer of responsible AI services - we are starting to use BBC R&D-developed image recognition, speech to text and content analysis tools to support our staff. We will lead the development of responsible personalisation by using data to create BBC services that are uniquely tailored to each of our users. We will seek ways to exploit the potential the BBC’s huge editorial and cultural assets through machine learning to enrich users’ lives. But we must do this in a way that upholds the public service values that have guided the BBC for almost a century, as set out above: impartiality, independence, accountability and universality. In other words, the BBC will be an exemplar for the responsible development of AI technologies in the interests of its audience. In practice, this means

¹³² <http://www.bbc.co.uk/news/av/world-asia-china-40073960/alphago-computer-defeat-painful-for-chinese-go-prodigy>

¹³³ <http://www.bbc.co.uk/programmes/p0571tdt>

¹³⁴ <http://www.bbc.co.uk/programmes/b08wm8zt>

¹³⁵ <http://www.bbc.co.uk/news/technology-40812697>

¹³⁶ <http://www.bbc.co.uk/programmes/b07dlxmj>

¹³⁷ <http://www.bbc.co.uk/programmes/b08wwnwkw>

¹³⁸ <http://www.bbc.co.uk/programmes/b08x9ckx>

ensuring that we use AI in a way which is free of commercial and political interests, is transparent in how it uses users' data and is equitable in the benefits it delivers to all sections of society. We know we will need to work hard to make sure that those who hold us to account - inside and outside the BBC, including our audiences - can truly interrogate our AI services, in the same way that our editorial processes allow for our journalism.

Conclusion

- 11 AI has significant positive potential, and we are confident this potential can be realised if AI is developed responsibly and in the public interest. We believe it's essential to give voice to our audiences through content to help them shape AI; we are committed to empowering them to contribute to the debate and the development of AI and to ensuring our own development of AI is compatible with the BBC's values and principles.
- 12 Nearly 100 years ago the BBC was transformed from a company to a corporation because there was recognition that nascent radio broadcasting was a truly transformative development, and necessitated the creation of a public service body that would put the needs of the public first. We believe AI has the potential to be even more transformative than radio, and in this moment, as was the case in 1929, leadership is needed to protect the public interest, and to shepherd the development of a mixed AI ecosystem for the greater public good. If we do not, then there is a risk that the gap between the winners and losers of rapid technological change will widen, and the positive potential of AI will be squandered. The successful introduction of AI - developed in the public interest - into society, will require like-minded public servants, commercial entities, government, and society more broadly to come together to collaborate. The transformative potential of AI makes this an imperative, and the BBC stands ready to play an active and positive role.

7 September 2017

BCS, The Chartered Institute for IT – Written evidence (AIC0049)

BCS, The Chartered Institute for IT response to the House of Lords Select Committee on Artificial Intelligence - Call for Evidence

1st September 2017

BCS, The Chartered Institute for IT

BCS is a charity with a Royal Charter. Its mission is to make IT better for society. It does this through leadership on societal and professional issues, working with communities and promoting excellence.

BCS brings together industry, academics, practitioners, educators and government to share knowledge, promote new thinking, educate, shape public policy and inform the public. This is achieved through and with a network of 75,000 members across the UK and internationally. BCS is funded through membership fees, through the delivery of a range of professional development tools for practitioners and employers, and as a leading IT qualification body, through a range of widely recognised professional and end-user qualifications. www.bcs.org

The pace of technological change

Question 1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1. The current state of artificial intelligence is that one specific set of technologies, “machine learning”, which has seen some impressive practical advances recently, is dominating, to the point where the two are practically synonymous in the public eye, and also where it seems to be difficult to secure funding for other forms of artificial intelligence research. As regards the challenges, and potential dangers, of machine learning, we refer to the Institute of Mathematics and its Applications evidence to this inquiry, and to the joint BCS/IMA evidence base at: <https://ima.org.uk/6910/the-debate-about-algorithms/> . We also note the IMA’s comments on how easy it seems to be to subvert the current machine learning systems, and feel that the risks are insufficiently appreciated, even by practitioners. We note the striking example they

give, quoted from [Evtimov *et al.*, 2017¹³⁹] of what clearly looks like a STOP sign, to human beings, but has been tampered with to appear like a 45 speed limit to a machine learning recogniser.

2. Stop sign to a human being, 45 speed limit to a machine [Evtimov *et al.*, 2017¹].



3. The committee, and indeed the public, needs to differentiate between Specific and General AI. Specific AI is where the intelligence is used to deal with a specific task, such as a medical diagnosis. In this case it seeks answers to pre-programme questions, but is incapable of forming new questions. General AI has the ability to ask new questions and is akin to human consciousness. Because we do not understand how human consciousness develops this is
4. Some way off (depending on what is meant by 'akin'), but will eventually raise ethical issues such as is the AI 'device' alive in the human sense. BCS wishes to caution society about the dangerous nature of trying to deploy analogues without full understanding of the ethical and sociological implications. A good example of the weaknesses of current general AI is shown by the Chinese insistence that two chatbots be taken offline for political incorrectness: <http://www.reuters.com/article/us-china-robots-idUSKBN1AK0G1> . A more humorous example of the difficulty of producing

¹³⁹ Evtimov *et al.*, 2017] Evtimov,I., Eykholt,K., Fernandes,E., Kohno,T., Li,B., Prakash,A., Rahmati,A. & Song,D., Robust Physical-World Attacks on Machine Learning Models. <https://arxiv.org/abs/1707.08945>.

even vaguely plausible General AI is given in <https://www.livescience.com/60275-ai-writes-next-game-of-thrones-novel.html> where the AI system produces text such as “Varys poisoned Daenerys and another of the dead men”.

5. Specific AI relies on human input to develop the questions, generally to identify “training data” and “testing data” which characterise the question, analyse the answers and define the action to be taken. Specific AI has the potential advantages of consistency, speed and not getting tired. Like other computer programs (because that is what it is), Specific AI can operate in real-time 24 x 7. It is likely to replace, or change significantly, many human jobs in the near (10 years) future which raises many questions about what roles humans will have in the future.

Question 2. Is the current level of excitement which surrounds artificial intelligence warranted?

1. A great deal of what is said in the popular media is, of course, wrong, utopian or alarmist. But nevertheless society is right to be deeply interested in major technological changes which have already affected (for a small example, see <http://metro.co.uk/2016/11/09/little-girl-uses-google-translate-to-invite-her-lonely-new-classmate-to-lunch-6246363/>), and will undoubtedly affect far more, nearly everyone’s lives.
2. Some of the developments in artificial intelligence, notably the developments in automated reasoning which allow the production and fielding of fault-free software in critical applications such as air traffic control, jet engine operation and medical devices (see IMA submission) are wholly beneficial. Others, notably those clustered around “machine learning” (which is better described as “machine pattern recognition”), are more mixed. If we look back at the history of the automobile, we see that we should avoid both the equivalent of the Locomotive Act 1865 (the “Red Flag Act”) and the equivalent of the period before driving tests

Impact on society

Question 3. How can the general public best be prepared for more widespread use of artificial intelligence?

BCS has no response to make to this question.

Question 4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

1. The impact of AI and Machine Learning is fuelling another generation of automation and is becoming an ever-stronger factor in the “appification” of

routine tasks, formerly done by the less skilled, on smartphones and tablets. Map software that anticipates your destination, navigates you there and offers you places to visit or stay is replacing work previously performed by personal assistants, travel agencies, and even friends. Tasks such as these and others in offices and factory floors that are routine or repetitive will be the first to be taken over by machines. Tasks that are complex are more likely, at least for the foreseeable future, to be assisted, rather than replaced, by machine intelligence. Most of the one fifth of activities in the U.S. workplace that McKinsey [2016¹⁴⁰] has identified as highly susceptible to automation are performed by low skilled or unskilled workers. Already changes due, at least partly, to technology are visible. The share of U.S. workers employed in routine office jobs declined from 25.5% to 21% between 1996 and 2015 [Economist, 2017¹⁴¹]. Far from being threatened by AI and Machine Learning highly skilled professionals such as doctors, lawyers and accountants will significantly benefit. Services in areas such as healthcare, law, and financial advice, are usually in high demand and can be made more affordable through AI. Professionals will be able to do more with less [e.g. Cohen, 2016¹⁴²] But in areas like retail and services, jobs are already being replaced by machines [e.g. Pierce, 2017¹⁴³].

2. The implications of these developments for policymakers are stark. First, policymakers need to understand that “unemployment caused by technology” is simply another way of saying “unemployment caused by a lack of skills”; second, when workers’ skills fall behind then inequality follows [Economist, 2017] In a highly educated and highly skilled society there is no reason for anyone to suffer unemployment due to displacement by technology for more than a short period. Other than managing worker displacement, for which Denmark’s flexicurity system
3. may offer a model, the best policy responses are likely to include measures for teaching people how to learn new skills. Workers who have learned how to learn are likely to need this skill throughout their careers

¹⁴⁰ McKinsey, 2016] McKinsey, ‘Where machines could replace humans—and where they can’t (yet)’, July, 2016, See: <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet>

¹⁴¹ [Economist, 2017] Lifelong learning: special report, *Economist*, January 14, 2017

¹⁴² [Cohen, 2016] Mark A. Cohen, 'How Artificial Intelligence Will Transform The Delivery Of Legal Services', *Forbes*, September 6, 2016, <http://www.forbes.com/sites/markcohen1/2016/09/06/artificial-intelligence-and-legal-delivery/20ebaa842647>

¹⁴³ [Pierce, 2017] David Pierce, 'This Robot Makes a Dang Good Latte', *Wired*, January 30, 2017, <https://www.wired.com/2017/01/cafex-robot-barista/>

Public perception

Question 5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

BCS has no response to make to this question.

Industry

Question 6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

1. It is very hard to say that a given sector cannot benefit from Artificial Intelligence. Take the thousands of people building and maintaining the Rolls Royce Trent 10000 engines, the only engines for the Boeing 787, Airbus A350-1000 and A330 neo. Their jobs are there only because of the small (less than 100) team who applied Automated Reasoning tools to verify the avionics and therefore give it that market edge. They themselves are only effective because of the team of about 10 who developed that methodology at Altran.
2. The barriers to adopting this sort of technology are largely demand-side ignorance, and supply-side lack of trained staff.

Question 7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

BCS has no response to make to this question.

Ethics

Question 8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

BCS has no response to make to this question

Question 9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

BCS has no response to make to this question.

The role of the Government

Question 10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

1. The EU's General Data Protection Regulation, due to enter in force in May 2018, talks, in Article 22(1), about "a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." At this point, the Regulation says (Article 14(2)(g), see also 15(1)(h)) that the subject must be "provided with meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."
2. While this seems plausible as a statement of general principles, it is extremely vague. Leaving it for clarification by case law seems unhelpful to the many practitioners. The meaning of "similarly significantly affects" is unclear. One significant question is whether this includes short-listing for a job.

Learning from others

Question 11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

1. In Answer 10 we noted that the EU's General Data Protection Regulation talking about "a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." As we said, this is plausibly a correct statement of general principles, it is extremely vague. The U.K. should learn from this vagueness, and seek, whether by legislation or by much firmer guidance from the Information Commissioners Office to give greater clarity to
2. "significantly affects" – this almost certainly includes the length of prison sentences, probably affects the granting of major loans (it does in the US: Fair Credit Act), but it is currently unclear whether this includes short-listing for jobs, university entrance etc.;
3. "meaningful information about the logic involved" – presumably this is more than "our deep neural net says you shouldn't be shortlisted", but how much more?
4. "based solely on" – there has been much concern in the US recently about "robo-signers", humans who sign papers prepared by computers, and do not effectively review them, and while one would hope that robo-signers were not sufficient to bypass "based solely on", the situation is not clear.

1 September 2017

Dr Simon Beard, Dr Seán Ó hÉigearthaigh, Dr Shahar Avin, Martina Kunz and Andrew Ware – Written evidence (AIC0150)

Executive Summary:

In the first section of this response, we focus on the benefits that artificial intelligence (AI) could bring and argue that not only the size, but also the distribution, of these benefits should be of primary concern to the Committee when weighing up the potential of developments in AI. This section primarily addresses the question “Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?”

In the second section, we focus on the potential for future advances in AI to pose catastrophic risk. This section primarily addresses the question “What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?”

In the third section, we focus on the governance challenge - how governments might intervene to mitigate risks and the potential for government intervention to make this problem worse as well as better. This section primarily addresses the question “What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?”

Section 1 – The benefits of AI, and their distribution

1. The potential benefits of AI are great, and include its potential to improve human welfare and to alleviate risks. However, since the risks associated with AI are likely to be equally spread across many people, or even concentrated on the worst off, it is not only the size, but also the distribution, of these benefits that must be considered.
2. AI has the potential to bring great benefits to the worst off in society. It could help overcoming implicit, institutional and persistent biases that often work against these people. It could remove educational and informational barriers to accessing economic, social, legal and cultural institutions. It could also reduce infrastructural barriers to technological development, because an AI system that is hosted in a technologically developed area can be easily accessed, via mobile technology, from around the world. Finally, it could help

- to solve pernicious global challenges such as how to allocate and manage resources, prevent conflict and respond to the violation of human rights¹⁴⁴.
3. On the other hand, many people emphasise the potential for AI to work against the interests of the worst off, by deskilling and automating industrial and service tasks, and thus removing opportunities associated with human and economic development¹⁴⁵ or by assisting with the surveillance and exploitation of the worst off. AI will likely increase the financial returns to capital relative to investment, which is associated with increased economic and social inequality.
 4. There is nothing inevitable about how the benefits of AI are distributed, both within the UK and around the world. It is worth remembering that the Industrial Revolution, the Technological Revolution (or Second Industrial Revolution) and the Digital Revolution (or Third Industrial Revolution) were equally responsible for creating highly unequal economies, such as China and the United States of America, and highly equal ones, such as Sweden and Japan. The same will almost certainly be true for the social and economic changes that will be brought about by the development of AI¹⁴⁶.
 5. One of the challenges in ensuring an equitable distribution of the benefits from AI is overcoming problematic bias¹⁴⁷. As algorithms use data that is historic, outputs reflect, and may even amplify, past injustices. In order for AI to produce more equitable outcomes it is therefore important that developers collaborate with the widest range of stakeholders and involve a broad range of perspectives. Due to unequal access to education and technology this is likely to require a proactive approach to encouraging more home-grown technological development in marginal communities and developing countries, something that the UK Government could play a valuable role in facilitating through the Department for International Development. Developers must also incorporate insights from those who will be using and affected by these technologies and cannot make the mistake, common in many previous technological developments, of assuming that the designer knows best. Further, the deployment of AI systems should respect existing socio-cultural structures, values, and practices; as well as future objectives of increased equality and sustainability.
 6. Another key challenge is the trustworthiness of AI systems. This will be critical to the acceptance of, and engagement with, potentially beneficial

¹⁴⁴ International Telecommunications Union (2017), *Accelerating the UN's Sustainable Development Goals through AI* <https://itu4u.wordpress.com/2017/03/27/accelerating-the-uns-sustainable-development-goals-through-ai/>

¹⁴⁵ Sutton Trust (2017), the state of social mobility in the UK https://www.suttontrust.com/wp-content/uploads/2017/07/BCGSocial-Mobility-report-full-version_WEB_FINAL-1.pdf

¹⁴⁶ World Economic Forum (2017) The Fourth Industrial Revolution: What it means, how to respond <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond>

¹⁴⁷ Not all forms of AI bias are problematic in this respect. See Danks, D. and London, A. L. (2017) Algorithmic Bias in Autonomous Systems <https://www.ijcai.org/proceedings/2017/0654.pdf>

technologies. It is challenging to accept guidance if it is not shown to be derived legitimately and credibly, especially when the foundational basis of the output of a system is not expressed explicitly – as is often the case with AI. Baroness Onora O’Neill has suggested that trustworthiness relies on three features – reliability, competence, and honesty¹⁴⁸. It seems evident that AI systems could be competent and reliable if deployed appropriately, but it is harder to see how an opaque system could be seen as truly ‘honest’, especially by people who lack specialised technical knowledge. It is important that AI has the feature of meaningful transparency, including not only the ability to explain how it makes decisions (offering evidence-based justification and natural-language explanations) but also by ensuring that people can find out what they want to know and what is important for them to know about these systems. This requires not only a high degree of explainability, but also that dissemination of appropriate information by developers and operators of AI systems.

7. It is, therefore, not enough to consider AI solely from the perspective of how the technology emerges, we must also consider the additional education and development interventions that will be needed to ensure that it is well used by people around the world.

Section 2 – AI and Catastrophic Risk

8. However, as a powerful technology with a broad range of applications, AI also poses a range of potential near- and longer-term risks as well as benefits. Some of the most important ethical implications for the development of AI relate to high impact risks that could be posed by future advances in AI over the long term, and especially the development of artificial general intelligence (AGI).
9. AI systems are currently able to perform at a level greater than or equivalent to that of human beings in some domains, such as chess, correctly answering trivia questions and stock market trading. However this performance is domain-specific. AGI is characterized by its ability to learn across any domain of knowledge or activity, to adapt to its environment and to exploit developments in one domain to make progress in another domain. If developed, AGI is therefore likely to be qualitatively very different to current AI¹⁴⁹. Progress towards AGI is currently at a very early stage. It is highly uncertain if and when general intelligence comparable to that possessed by humans might be achieved in artificial systems, although it is highly unlikely to be achieved within the coming decade.

¹⁴⁸ O’Neill, O. (2015) Trust, Trustworthiness and Transparency <http://www.efc.be/human-rights-citizenship-democracy/trust-trustworthiness-transparency/>

¹⁴⁹ Goertzel, B., Hitzler, P., & Hutter, M. Artificial General Intelligence <http://www.hutter1.net/ai/agifb09.pdf>

10. The probability of AI systems posing catastrophic risks is highly uncertain, and may be low. Nevertheless a number of recent reports,^{150,151} scientific editorials¹⁵² and policy discussions¹⁵³ have highlighted that they may indeed pose such risks. Due to their rapidly developing nature, such risks are likely to be understudied, and may not be adequately evaluated in global policy and technology governance analyses. In his 2014 Annual Report the Government Chief Scientific Advisor argued that such risks should not lead us to abandon the development of new technologies like AI, but that “we must constantly scan the horizon and do our best to prevent and mitigate adverse consequences of new technologies”¹⁵⁴.
11. If powerful AGI systems are achieved in the future, they are likely to be able to solve problems and manipulate their environment to an extent equivalent or greater to that of humans. One plausible scenario (being explored by several research groups) of how they could pose catastrophic risks would be if an AGI system possessed the following properties:
 - Goal non-alignment: the AGI system’s goals were not sufficiently well-specified to avoid the possibility of catastrophic consequences and
 - Decisive advantage: the AGI system was sufficiently capable and unconstrained in its operations that anticipating, preventing or modifying its activities was difficult or impossible. An AGI system might gain such an advantage by achieving a qualitatively higher level of problem-solving ability than humans, operating at significantly faster speeds than humans or being significantly better at coordinating its activity on a global scale than humans¹⁵⁵.
12. Other circumstances under which a future AI or AGI system could present a catastrophic risk are if it gave a decisive advantage to an individual or group who misused it, if its creation precipitated a catastrophic event such as a global war or if, despite its advanced capabilities, it still made imperfect decisions, leading to its actions precipitating a catastrophe in ways similar to

¹⁵⁰ World Economic Forum Global Risk Report 2015

http://www3.weforum.org/docs/WEF_Global_Risks_2015_Report15.pdf

¹⁵¹ Cotton-Barrett (2016) Global Catastrophic Risks 2016

<http://www.globalprioritiesproject.org/wp-content/uploads/2016/04/Global-Catastrophic-Risk-Annual-Report-2016-FINAL.pdf>

¹⁵² Rees, Martin. "Denial of catastrophic risks." Science (2013)

<http://science.sciencemag.org/content/sci/339/6124/1123.full.pdf>

¹⁵³ United Nations CBRN National Action Plans: Rising to the Challenges of International Security and the Emergence of Artificial Intelligence

http://un.mfa.gov.ge/index.php?lang_id=ENG&sec_id=149&info_id=33437

¹⁵⁴ Peplow, M. (2014). Innovation: managing risk, not avoiding it

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/381905/14-1190a-innovation-managing-risk-report.pdf

¹⁵⁵ Price, H. (2012). Artificial Intelligence - can we keep it in the box?

<https://theconversation.com/artificial-intelligence-can-we-keep-it-in-the-box-8541> Stuart Russell, S. (2014) Of Myths and Moonshine <https://www.edge.org/conversation/jaron-ianier-the-myth-of-ai>.

those in which human activities have resulted in climate change and biodiversity loss.

13. The consensus amongst experts is that AGI is likely to have significant benefits to humanity, but that there is some risk of producing 'extremely bad' or 'catastrophic' outcomes¹⁵⁶. Furthermore, whilst many researchers are confident that AGI with greater than human capabilities can be developed in theory, most believe that it is still some way away. In a recent survey of experts across a number of fields related to the development of AI, median estimates were that there is only a 10% chance of having developed this level of AGI by 2026, but a 50% chance of having developed it by 2062¹⁵⁷. However, these results cover a very wide range of opinions, with some experts arguing that AGI, if it is even possible, will take centuries to develop.
14. A growing community of researchers are working to understand the long-term risks posed by AI, with UK researchers playing a leading role. As well as the Centre for the Study of Existential Risk, significant research projects in the UK are being pursued at the Leverhulme Centre for the Future of Intelligence (at the Universities of Cambridge, Oxford, Imperial and Berkeley), and the Future of Humanity Institute (at the University of Oxford).

Section 3 - AI governance

15. The challenges of avoiding catastrophic risk from the long-term development of AI are complex, and given the likely time-frame it would be unwise to reject the near term benefits of AI because of them. However, it is important that the avoidance of catastrophic risk forms part of the discussion about the ethical implications of the long-term development of AI. At present, research into policy to avoid catastrophic risks from AI is focusing on two key areas:
16. Firstly, how to effectively combine efficient governance of long term and short-term risks from AI, so that as many people as possible can enjoy the benefits of AI. As the White House Office for Science and Technology Policy noted: "the best way to build capacity for addressing the longer-term speculative risks is to attack the less extreme risks already seen today, such as current security, privacy, and safety risks, while investing in research on longer-term capabilities and how their challenges might be managed... Although prudence dictates some attention to the possibility that harmful superintelligence might someday become possible, these concerns should not be the main driver of public policy for AI."¹⁵⁸

¹⁵⁶ Bostrom, N., & Cirkovic, M. M. (Eds.). (2011). *Global catastrophic risks*. Oxford University Press. <http://global-catastrophic-risks.com/docs/global-catastrophic-risks.pdf>

¹⁵⁷ Grace, K. et al (2017), When will AI exceed human performance?, <https://arxiv.org/pdf/1705.08807.pdf>

¹⁵⁸ White House Office for Science and Technology Policy (2017) https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

17. Secondly, what lessons can be learned from regulating other dual use technologies that have long term benefits but pose potentially catastrophic risks, such as nuclear power and biotechnology. As with these industries, it may prove challenging to monitor and enforce AI governance, particularly if the necessary computing hardware is widely available or easily acquired, and if access to the software and necessary knowledge is difficult to control.
18. Whilst there are many near term challenges in developing AI that deserve a hands on approach to governance and regulation, for developments that are only likely to occur considerably in the future, such as AGI, it is currently very difficult to recommend concrete policies to guide progress at this point; too many uncertainties exist both regarding the specifics of the technology, and the context within which it may be developed. However our research implies 4 key conclusions:
19. Firstly, too much competition in the development of AI and its applications, either between different research groups or different countries, might pose risks, particularly after certain thresholds in capability have been achieved. While we are still a considerable distance from risky thresholds in AI development, processes that encourage collaboration ahead of time between leading research groups on near- and long-term safety relevant issues should be encouraged. Inspiration might be drawn from the 'precompetitive sharing' of safety relevant information in the aerospace and automotive industries. One near term example is that it would be valuable if self-driving car technology companies shared data on driving accidents and near-misses.
20. Secondly, it is important that research into AI safety be an integral part of research into AI development. We are enthusiastic about the potential to develop strong industry norms around AI safety and the societally responsible application of AI through the work of organisations such as the Institute of Electrical and Electronics Engineers (IEEE)¹⁵⁹.
21. Thirdly, whilst it is hard to anticipate exactly what problems will emerge in the future for AI safety research, it is important the research is encouraged right now in order to build capacity and solve what problems can be addressed in the present day. An important example of this is technical work, involving collaborations between AI safety researchers and developers, on research agendas such as the 'Concrete Problems in AI Safety' paper¹⁶⁰ and the safe behaviour of reinforcement learning agents¹⁶¹.
22. Finally, whilst governments can support the safe and beneficial development of AI, there are also risks associated with government action, and regulation of general research and development in AI is likely to be particularly difficult to design and implement effectively. For instance, regulation may push AI development away from responsible researchers in well regulated

¹⁵⁹ see IEEE (2016) Ethically Aligned Design
http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf

¹⁶⁰ Amodei et al (2016) Concrete Problems in AI Safety <https://arxiv.org/abs/1606.06565>

¹⁶¹ Orseau, L. and Armstrong, S. (2016) Safely Interruptible Agents
<https://intelligence.org/files/Interruptibility.pdf>

environments and towards irresponsible individual developers and less well regulated regimes. CSER researchers have identified a number of ways in which regulation may have such undesirable effects. These depend on the cause of this failure (incorrect analysis of the risks posed by AI, poor design of the regulatory response or lack of coordination between different actors), the mechanism by which it occurs (regulation of the wrong thing, in the wrong way or at the wrong time) and the kind of bad outcome that is produced (risky AI is allowed to be developed, non-risky and beneficial AI is prevented from being developed or perverse incentives are created for AI developers that have negative consequences).

23. We therefore recommend that the Committee consider how the UK Government can build institutions that will draw on the experience of diverse stakeholders to ensure a governance environment that can respond dynamically to the development of AI. For example, establishing a standing Commission on Artificial Intelligence, as suggested by the Academic Director of the Centre for the Study of Existential Risk, Professor Huw Price, and others and endorsed by the Science and Technology Committee's report on robotics and intelligence. We also recommend limiting potential regulatory activities to sector-specific applications of AI, in collaboration with sector-specific governance bodies, at this time.

Note

This response was written with additional input from Dr Seán Ó hÉigearthaigh, Dr Shahar Avin and Haydn Belfield of the Centre for the Study of Existential Risk, Martina Kunz at the Centre for the Future of Intelligence and Andrew Ware of the University of New Hampshire. The Centre for the Study of Existential Risk is an interdisciplinary research centre within the University of Cambridge dedicated to the study and mitigation of risks that could lead to human extinction or civilisational collapse. Dr Simon Beard is part of the 'Managing Extreme Technological Risks' project, funded by a grant from the Templeton World Charity Foundation. The opinions expressed in this response are those of the authors and do not necessarily reflect the views of Templeton World Charity Foundation.

6 September 2017

Miles Berry – Supplementary written evidence (AIC0247)

Artificial Intelligence, SEND and the Computing Curriculum

Miles Berry, University of Roehampton

Artificial Intelligence and SEND

There are a number of ways in which limited AI can help pupils with special educational needs and disabilities (SEND) access school, and few would be surprised if future developments did not result in further affordances to support inclusion and accessibility.

The 'expert system' approach to AI, and more recent work in machine learning, has been applied with a degree of success to diagnosis of some special educational needs (SEN), including specific language impairment, attention deficit, and dyslexia and related difficulties. At least one machine learning based tutoring system appears able to integrate dyslexia identification into routine activity presentation and assessment. Such approaches might allow schools to identify children for professional diagnosis by educational psychologists more reliably than simply using teachers' or parents' judgement.

Text to speech applications, although typically rule based rather than powered by AI, can make it possible for visually impaired pupils to access a far broader range of texts than would formerly have been possible, and can also be of benefit to other children for whom reading is a challenge, such as those with dyslexia. More recent applications of AI allow images, including live camera feeds, to be described in text, or speech.

Speech to text has advanced rapidly in recent years, and many are now familiar with tools such as Siri, Alexa and Google Assistant. Automatic transcription of spoken language is increasingly accurate: hearing impaired pupils can access YouTube videos through automated captions, and a live transcript of a teacher's introduction to a topic can be made available to hearing impaired pupils and those who would benefit from being able to look back over what the teacher had said. Similarly, pupils themselves can use speech recognition to 'type' answers to questions, stories or essays; this can also be of benefit for pupils with dyslexia or visual impairment. The corpus on which these systems are trained include relatively little speech by young children, and thus results are currently rather less accurate than for adult speech.

Whilst not a SEN, pupils for whom English is a second language can use fully automatic machine translation to follow instructions and access lesson content and texts. It also allows them to participate in the lesson and to ask questions of their teacher. Image and speech recognition can be used to provide translation between sign language and written or spoken text, although most of this work

has been conducted in American Sign Language rather than British Sign Language at present.

Tools as simple as spelling and grammar checkers can be used by all pupils to correct errors in their work, but may be particularly useful for pupils with dyslexia: it's not clear whether such tools help pupils to *learn* correct spelling and grammar, but it's possible that some application of machine learning alongside the data generated by these tools would help. Pupils, including those with dyslexia, may also benefit from AI based text simplification tools in order to access complex texts.

AI based automatic, personalised tutoring systems might be particularly useful for pupils with SEN, where the usual path through a course of study may not be ideal - with enough prior data, machine learning algorithms might well be adept at tailoring a sequence of learning activities more appropriately for an individual learner than a teacher would be. Such tools may be particularly helpful for pupils with attention deficit or autism spectrum disorders who might find the demands of a typical classroom environment uncondusive to study. By analogy with GPS-based navigation software: not all journeys start from the same place, and not all traffic can follow the same route, even if the destination is the same.

AI has been used to support pupils with Autism spectrum disorders (ASD): human-like robots can be programmed to behave in a predictable way, making interactions less threatening for pupils with ASD, and helping them develop some mental model of how the robot will react, perhaps making it a little easier to construct a theory of mind for human interaction. There are reports of children with ASD developing conversation skills through interaction with Siri, and subsequently applying these to interactions with people.

SEND and the computing curriculum

As computing is a national curriculum subject, pupils with SEND have the same entitlement to be taught the curriculum content as any other pupils: schools have an obligation to ensure that there are no barriers to pupils' attainment, and that specialist equipment and different approaches are provided where these are needed. Operating system developers have done much to build accessibility in to their products in recent years and pupils should be taught how to make effective use of these features.

In computer science lessons, particular attention needs to be paid to the programming environments used by pupils. The popular Scratch block based language is at present inaccessible to visually impaired pupils or those unable to use a mouse, trackpad or touch screen; alternatives which work with screen-readers and keyboard input are available. Similarly, whilst block-based languages such as Scratch are generally more accessible than traditional text-based languages, even Scratch places significant demands on pupils' reading abilities: simpler, icon-based, alternatives are available. On the other hand,

Scratch does provide excellent support for pupils to program in languages other than English.

An inclusive approach to computing should ensure an appropriate balance between the foundation (computer science), application (information technology) and implication (digital literacy) elements of the curriculum. For some pupils with SEND, too great a focus on programming and other aspects of computer science at the expense of IT skills and online-safety may do little to prepare them for the practical needs of their subsequent study, employment and adult life. Particular attention should be paid to ensuring that pupils who are more vulnerable because of SEND have a secure understanding of how to keep themselves safe, and of their responsibilities, when using the internet.

The acceptance of the Rochford Review recommendation that the former P-scales be removed for assessment in computing below the level of the national curriculum has left a gap between assessment for those not engaged in subject-specific learning against aspects of cognition and learning and assessment in accordance with the national curriculum attainment targets based on the programmes of study. Schools need guidance in teaching and assessing the progress of those pupils not yet working at the level of the national curriculum, and whilst this is available for English and mathematics, the removal of the P-scales impedes this for computing and other foundation subjects.

The computing national curriculum provides an opportunity for pupils to think about user-centred design, including accessibility and inclusion, when developing their own programs and other digital content. In learning about effective design pupils should be taught that good design is inclusive. In practice, this can be as simple as providing both spoken and textual instructions in a game of their own design, or the addition of subtitles to a video they edit, but at a higher level might include the design, development and testing of products designed to support users with particular needs, impairments or disabilities.

A substantial body of work suggests that software engineering employs a disproportionately high number of adults with ASD. For pupils with ASD, the opportunity to learn to program whilst at school may provide greater confidence and sense of achievement, as well as providing a path on to further study and employment in this area. It is important that schools allow pupils, including those with ASD, who express an interest in or aptitude for programming to study GCSE and A Level computer science, even if they might fail to meet the school's normal requirements for courses of this rigour.

17 January 2018

Big Brother Watch – Written evidence (AIC0154)

About Big Brother Watch

Big Brother Watch is a civil liberties and privacy campaign group. We campaign to give individuals more control over their personal data, and hold to account those who fail to respect our privacy, whether private companies, government departments or local authorities. We have produced unique research exposing the erosion of civil liberties in the UK, looking at the dramatic expansion of surveillance powers, the growth of the database state and the misuse of personal information.

INTRODUCTION:

Artificial intelligence (AI) is becoming an unavoidable element of 21st Century life.

AI currently takes many forms including search engines, voice recognition, product or service recommendation systems, photographic analysis and recognition, targeted advertising, and virtual assistants such as Apple's Siri, Microsoft's Cortana, Amazon's Alexa and Google Home.

AI influences the products we purchase, the news we read, the adverts we see and potentially who we vote for. AI is also becoming crucial to the functioning of the economy, being used to carry out trades, decide credit scores, and calculate and decide on financing and lending.

Large scale datasets are the fuel for many, if not all AI initiatives. The large scale acquisition, retention and use of both industrial and personal data brings privacy, security and data protection issues to the fore. This is particularly so if AI is to be used to simulate human decision-making, at which point the very serious problems of biased and prejudiced AI must also be raised.

DEFINITION OF 'ARTIFICIAL INTELLIGENCE (AI)'

We have followed a wide interpretation of AI, including machine learning, which concerns the imitation of human intelligence in an artificial manner, by computer programs, systems or algorithms. This technology can be used to analyse data and make decisions in a similar way to a human.

RESPONSE:

IMPACT ON SOCIETY

Question 3: How can the general public best be prepared for more widespread use of artificial intelligence?

Prepare the public by increasing their understanding and engagement

1. AI is already around us, making important decisions for and about people. However, alarmingly, most people are unaware of what AI is and how it works. This clearly needs to change, but we believe there is a need to go back to basics and engage in educating people about what their data is, as well as

the value and importance of their data. Following this, explanation of AI can then flow naturally; the public will understand the fundamental issue that personal and commercial data will power AI, that such data is generated by people and that it can impact how people live.

2. AI, like data, is invisible. It runs in the background of online services like Amazon or Google, so the public are unable to see how it is used, what it is used for and what the benefits or potential harms are. This leads people to be generally ignorant of what AI is and the extent to which it is currently used.¹⁶² Most people will be unaware that the helpful recommendations they get when they visit a website are created using AI. Most people are also unaware that it is their personal data that fuels AI. It would be helpful for the public to understand that a “smart” product like one powered by AI does not start smart – it only becomes smart because it is trained on information that we give to it. The more it learns about us, the smarter it becomes, but obviously that requires us to tell it everything we can and provide it with large amounts of personal data.
3. The General Data Protection Regulation, in the form of the new Data Protection Bill, will help these conversations to take place, particularly in relation to get people to engage with their rights and responsibilities when it comes to data. But further work can and indeed must be done by Government to alter their current approach as demonstrated in Part 5 of the Digital Economy Act of keeping people at arm’s length from their data, to ensuring and encouraging people the right of control over how their data is used.

PUBLIC PERCEPTION

Question 5: Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

Public awareness of AI

4. The public must be informed of the affects and effects of AI on their privacy, their security and their data protection. The public know they have an ‘online footprint’ but rarely understand how they can control access to that data and prevent it from being used for AI purposes or for other purposes than those they can easily see and control. Educational work needs to be given priority to

¹⁶² The Royal Society, *Machine learning: the power and promise of computers that learn by example* (April 2017), pg85.

([https://royalsociety.org/~media/policy/projects/machine-](https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf)

[learning/publications/machine-learning-report.pdf](https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf)) Last accessed 24/08/2017.

ensure people understand their role and duty of responsibility as a digital citizen.

Transparency and Interpretability of AI

5. Whilst the corporate concerns regarding the intellectual property of AI and algorithms are well known, we must be very careful not to place greater emphasis on, and protection of, the rights of corporations, over the rights of the citizen whose data is being used to fuel the products and services offered by the public and private sectors.
6. Because AI can fundamentally impact a person's life, moves should be undertaken to ensure that the transparency of AI programs is standard, particularly when AI is used to make a decision affecting people or impact how people live their lives. The public must always be fully aware of when they are subject to, or affected or impacted by a decision made by AI. Increased transparency and accountability of public-facing AI, including the methods behind the system, and the reasons for decisions, will not only benefit society as a whole in terms of open source information but will increase public trust and confidence and subsequently, public engagement with AI systems.
7. We welcome the moves in the General Data Protection Regulation (GDPR) towards greater protection over automated machine-based decision-making and profiling, however the protections offered are just the start: as AI and interconnected technologies take greater hold, work to ensure the protection of people's digital lives will need to be monitored closely.

Privacy and cyber-security

8. Promoting the fundamental importance of protecting privacy in AI systems, such as the 'privacy by design' approach, should be encouraged as an industry standard. If the public are confident that the systems they use are protecting their private information, and that they themselves are in control, their confidence in the technology will improve. Privacy by design and security by design are well established concepts encouraged to ensure that from the very beginning, during the research and development of a project, the security and privacy of data, of the system, and of the individual using the system are built into the design, and not left as an afterthought at the end of production.

9. The impact of cybercrime is reported to cost the UK billions of pounds a year and is only set to grow as a problem.¹⁶³ It is therefore no longer acceptable for companies to sell devices which offer little to no protection for citizens and organisations alike. It is also not acceptable for public or private sector organisations to adopt technologies including AI without ensuring cyber-security protections are standard.
10. The encryption argument is contentious and one which is often presented, disingenuously, as a zero sum game. Whilst solutions for societal threats are consistently being sought, we must be careful not to undermine the security of all in a connected society as a reaction to other threats. The importance of end-to-end encryption as a much needed and fundamental tool for the security of the digital citizen in a digital world, which protects the security of more people than it harms, as well as organisations and national infrastructures, must be recognised and championed.

AI and the democratic process

11. The use of AI-driven analytics in relation to the democratic process is a growing area of concern. Analysis of people's data in order to determine how they may vote and what their specific areas of concern are nothing new, but the connectivity of communications, the impact of social media as a platform for sharing ideas and the ability to harvest, analyse and make conclusions from that data is a capability which is only now being realised, due to the capabilities of AI.
12. AI can analyse publicly available information people have posted online and draw personal insights from it.¹⁶⁴ Basic public datasets, such as Facebook 'likes', can be analysed to make predictions on people's political views.¹⁶⁵ This information is clearly of real value to political campaigns during elections.
13. There are restrictions on how much money can be spent by political parties during elections on campaigns, including online campaigns. However, there is an issue with the rise of 'dark advertisements' – an election-related message, targeted at a specific group or groups based on such publicly available information. The prevalence of these 'dark ads' during the 2017 General election was documented by the website 'Who Targets Me'.¹⁶⁶

¹⁶³ National Crime Agency (2016) <http://www.nationalcrimeagency.gov.uk/publications/709-cyber-crime-assessment-2016/file>

¹⁶⁴ The Royal Society (2017), *Machine learning: the power and promise of computers that learn by example* (April 2017), pg90

¹⁶⁵ Kosinski M, Stilwell D, Graepel T (2013), *Private traits and attributes are predictable from digital records of human behaviours*, PNAS 110 5802-5805

¹⁶⁶ Who Targets Me website: <https://whotargets.me/en/>

14. As with the use of traditional media in political campaigns, such data use for political purposes must be scrutinised. We welcome the investigation being undertaken by the Information Commissioner's Office into the use of data analytics for political purposes, but see this as just the start.

ETHICS

Question 8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

AI exhibiting and reinforcing bias

15. Data and the algorithms they populate are meant to be free from prejudice and bias. However it is well reported that bias and prejudice within data which is used to train AI can lead to bias and prejudice in the results.¹⁶⁷

16. This happens in all aspects of AI, from advertising to insurance to healthcare. We have chosen to draw your attention to the problems of bias in AI in relation to policing and criminal justice.

17. For example, a US court computer program – Compas – designed to assess the risk of re-offending, was discovered to have “turned up significant racial disparities”; the algorithm “was particularly likely to false flag black defendants as future criminals, wrongly labelling them this way at almost twice the rate as white defendants”.¹⁶⁸

18. It was reported in May 2017 that Durham Police are preparing to use AI to decide whether, like the US Compas system, suspects should be kept in custody.¹⁶⁹ The system uses data beyond a suspect's offending history, including their postcode and their gender¹⁷⁰ to assess the risk of re-offending, and contributes to the decision whether to keep a suspect in custody or release them.

19. Systems such as those used by Durham police, facial biometric technology – another form of AI – and issues relating to bias and false positives being

¹⁶⁷ The Guardian (2017) <https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals>

¹⁶⁸ ProPublica (2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

¹⁶⁹ BBC News (2017) <http://www.bbc.co.uk/news/technology-39857645>

¹⁷⁰ *Ibid*

determined by poorly built algorithms, are a very real concern. Extensive research has been undertaken in the US outlining problems of AI in this area.¹⁷¹ We draw this to your attention as there are increased moves to roll out facial biometric systems by UK police forces with vast investment coming from the Home Office. There has been no parliamentary scrutiny of these plans yet we know such technologies have been used at the Champions League Final in Cardiff this year, Notting Hill Carnival in 2016 and 2017 and Download Festival in Leicester in 2015. Furthermore we know that police forces are building their own facial biometric systems which are being used to make algorithms of people who are un-convicted of any crime or wrongdoing – challenging the concept of innocent until proven guilty.

20. Biased AI is an extremely serious concern in relation to fundamental civil liberties of equality and non-discrimination. Any such automated decision making must be subject to regulation and oversight. Intrusive surveillance technologies are consistently being purchased and rolled out by law enforcement, local authorities and official organisations without any debate in parliament, or any regulation or legislation. This is a very worrying trend, particularly when the technology is being trialled when its abilities are far from accurate.

AI and Privacy

21. It is an ongoing concern that the more data you have the better the outcomes are. This is a misnomer: the quality of data is critical, not the quantity. Citizens consistently raise concern about the lack of control over the data they are asked to provide, in order to access or benefit from a service; we have seen this from our own research.¹⁷² As AI grows the need for more and more data in order to support the system's ability to learn, grow, and, subsequently apply learning will be phenomenal. As a result, there is the potential that vast amounts of sensitive and or personally identifiable data will be collected, such as being 'scraped' from the internet. This is a huge concern for people's personal privacy.

22. Whilst education will be critical, a different approach must be encouraged by Government. For example, the Digital Economy Act uses the word "wellbeing" as a reason for bulk data sharing. "Wellbeing" is ill-defined and has been heavily criticised by the Supreme Court during the Scottish Parliament's controversial Named Person Scheme which intended on sharing the data of all Scottish children with a non-adult parent in order to protect them. The

¹⁷¹ The Atlantic (2016) <https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/>

¹⁷² Big Brother Watch (2015) <https://www.bigbrotherwatch.org.uk/wp-content/uploads/2015/03/Big-Brother-Watch-Polling-Results.pdf>

overarching view was that “wellbeing” falls short of the standard data protection view that data should only be used when “vital”. Such nudge tactics in the area of data, privacy and AI are worrying and must be addressed.

Anonymisation

23. The analytical capabilities of AI in the ‘big data’ environment have the potential to completely undermine formed notions of privacy, especially in the context of ‘anonymised’ data. We have consistently raised concern about the promises of anonymisation as a panacea.

24. There are countless studies where researchers have re-identified people from anonymised datasets. We would recommend the committee look at the work of Professor Latanya Sweeney PhD (Professor of Government and Technology at Harvard University, Director of the Data Privacy Lab and former Chief Technologist at the Federal Trade Commission) who proved that 100% re-identification was possible even when the data was anonymised.¹⁷³ By taking the South Korean Resident Registration Number – which closely matches the makeup of the UK’s NHS number – Professor Sweeney was able to re-identify all citizens using two entirely different methods.

Consent

25. We welcome the moves in the GDPR to improve the way citizens are required to give consent to how their data is used and what protections organisations must undertake to ensure informed consent has been given. However there remains little requirement for organisations to fully inform people of who their data might be shared with and for what specific purpose. This remains a very serious concern. If citizens’ data is to become part of the product – as we see with many AI technologies – there should be far greater transparency of how data will be acquired, used, shared and stored, with specific informed consent to be given and withdrawn if necessary, with no detriment to the individual.

Question 9: In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable? When should it not be permissible?

Transparency and interpretability of AI

26. Certain forms of AI such as neural networks or ‘black-box’ systems can be virtually impossible to audit because of their very nature: a process which is hidden and always changing. This can result in a serious accountability deficit. An example of this is the risk of re-offending algorithm used by Durham

¹⁷³ Technology Science (2015), De-anonymizing South Korean Registration Numbers Shared in Prescription Data, 29th September 2015: <https://techscience.org/a/2015092901/>

Police. If the decision making process is unknown and cannot be analysed, this precludes fundamental and basic principles of oversight and accountability, and runs the risk of limiting a judge's ability to render a fully informed decision.¹⁷⁴

27.If we don't fully understand the AI we create or whose decisions we are subject to, we won't be able to predict or pre-empt failures, and we won't be able to address their failings. If an individual is subject to a decision by AI, but is not able to know the reasons for the decision, or the decision-making process, this results in an unacceptable accountability deficit.

28.Automated decision-making systems are already in use and are governed by Section 12 of the current Data Protection Act. We are pleased to see the protections emphasised in the GDPR with what is effectively a challenge to the "computer says no" approach to decision making and the encouragement of a human point of view as a right. However, the more widespread use of more advanced AI programs must also be subject to the same regulation, and there must not be loopholes. In the same way that a public body must be publicly accountable, where AI programs involve, affect or impact the public, they too must be accountable, but they must also be transparent; the programming of AI and its inner-workings must be open for scrutiny.

THE ROLE OF THE GOVERNMENT

Question 10: What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

29.Artificial Intelligence clearly has the potential to be immensely powerful and provide a wealth of benefits to individuals and society as a whole, but for the benefits to be achieved protections will need to be put in place to ensure that individuals are not put at risk by machine learning, algorithmic bias or poor data protection.

30.Government has the opportunity to lead the way in establishing a new approach to how we live in a connected society, as opposed to falling in line with the approach taken by big business.

31.We would like to see independent oversight of AI in the form of a regulatory or supervisory body to provide legal and technical scrutiny of AI technology and algorithms.

¹⁷⁴ Wired (2017) https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/?intcid=inline_amp

32. With regard to the use of AI technologies for policing or in the criminal justice system, any system which uses machine learning, AI or algorithms to police society must be subject to independent scrutiny and parliamentary debate before it is implemented.

6 September 2017

Big Innovation Centre – Written evidence (AIC0119)

1. Introduction

Big Innovation Centre - established in 2011 – is working to build a global innovation and investment hub by 2025, create great companies and make the world more purposeful and inclusive through the enormous potential of technology, creativity and innovation.

As part of its core activities, **Big Innovation Centre** has been **appointed Secretariat for the All-Party Parliamentary Group on Artificial Intelligence (APPG AI)** launched in January 2017. The group aims to explore the impact and implications of Artificial Intelligence, including Machine Learning. Prior to this, we have been active in the field of AI through different initiatives such as big data project with our corporate partners and public agents.

Big Innovation Centre is also (i) leading a reporting and valuation project on the intangible economy, entitled Intangible Gold (working group includes researchers from Bank of England, Office of National Statistics and Oxford University) and (ii) building and piloting a digital platform and AI analytics tools to audit the UK economy and innovation supply system (entitled The National Innovation Audit).

In response to the Select Committee’s call for evidence, Big Innovation Centre will be using the evidence gathered from ongoing and recent projects, including:

Evidence	Big Innovation Centre projects	Documentation
Evidence 1	APPG AI: Four High-Level Parliamentary Meetings to date (i) to unpack the term 'Artificial Intelligence'; (ii) to gather evidence to understand it better; (iii) to assess its impact; and,	<ul style="list-style-type: none"> • Theme Report 1: What is AI? (2017) • Theme Report 2: Ethics and Legal in AI: Decision Making and Moral Issues (2017) • Theme Reports 3: Ethics and Legal: Data Capitalism (Forthcoming 2017) • Theme Report 4: Markets and AI-Enabled Business Models (Forthcoming 2017) <p>All reports and minutes from meetings can be downloaded from the APPG AI</p>

	ultimately (iv) to empower decision-makers to make policies in the sphere.	web site: www.appg-ai.org
Evidence 2	Intangible Gold research programme	<ul style="list-style-type: none"> • Intangible Asset Reporting and an Intangible Assets Charter (2017) • Intangible Asset Reporting: Defining Britain’s Real Treasures 2017 <p>Both reports can be downloaded from BIC web site: http://biginnovationcentre.com/publications</p>
Evidence 3	The Future of Trade Think-Piece - prepared for Innovate UK in July 2017, to understand ‘who, what, where & how’ automation and Artificial Intelligence are disrupting the marketplace	<ul style="list-style-type: none"> • The Future of Trade (2017) <p>Report can be downloaded from BIC web site : http://biginnovationcentre.com/publications</p>
Evidence 4	The National Innovation Audit – a project under development and implementation, aiming to illustrate the UK innovation ecosystem	<ul style="list-style-type: none"> • Pilot work described in background document for The Innovators Board (January 2017) <p>Report can be downloaded from BIC web site: http://biginnovationcentre.com/publications</p>

	through an online platforms and advanced data analytics.	
Evidence 5	Big Innovation Centre big data report and hack day with Camden Council featuring	<ul style="list-style-type: none"> Lessons Learnt From a Hackathon (2013) <p>Report can be downloaded from BIC web site: http://biginnovationcentre.com/publications</p> <p>Guardian features from 2013: first “Councils call in the geeks to help them solve local problems” and later “Big data: Camden Council leads the digital revolution”).</p>

AI Definition (see Evidence 1 and Evidence 3):

To respond to the questions posed by the Select Committee, we are adopting the more broad, general definition of Artificial Intelligence (AI). Hence, from this point onwards, we will be referring to AI as an umbrella term to describe several advances in technology (as opposed to market products), in fields such as Machine Learning, Deep Learning, robotics, autonomous decision-making, natural language understanding, and neural networking.

Furthermore, we will be focusing on the implications of Narrow AI (otherwise known as weak or non-sentient AI) rather than General AI. Our evidence shows that most advances happening now and in the short-term horizon are examples of the former category, a type of AI that is successful at performing a single task but unable to understand and reason with the environment as a human would. For point of reference, in our *Theme Report 1: What is AI?*, we have included excerpts from thought leaders in the space explaining what AI is to them.

2. Industry - How can the data-based monopolies of some large corporations, and the ‘winnertakes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy? (Q7)

Big Innovation Centre’s vision for a Data Charter (see Evidence 2 and evidence 5):

Background: Opportunities for public services as incentive for big data sharing

Big Innovation Centre co-hosted an all-day hackathon with Camden Council, big corporates and computer scientists in 2013. It was in relation to how big data could make public services more efficient with respect to repair services on council housing, crime and the ambulance service, and featured in the Guardian at the time (“Councils call in the geeks to help them solve local problems”¹⁷⁵ and later “Big data: Camden Council leads the digital revolution”).¹⁷⁶

We were also inspired by the New York City Mayor’s Geek Squad that opened-up the archives of their boilers and sprinkler systems, the state of their local taxes, the number of heart attacks and fires that occur inside their buildings and whether they have ever logged complaints about roaches or construction noise. Additional data was gathered about their businesses, their commuting habits and more including parking meters and those receiving tickets, and much more.

We found that despite the obvious opportunities of big data (public data!) they could not be opened to combat the investigation in sufficient details, and online public data access and text mining from the web or other AI tools can only get us so far. Worse, data was not reported on central topics, e.g. in relation to social housing. Three to four years on, nothing, or very little, have changed, despite the Shakespeare Review and the subsequent government’s data strategy citing our work at the time. With the recent Grenfell tower tragedy, the situation has become much more urgent.

The solution: Big Innovation Centre proposes a ‘Data Charter’ on the uses of personal and business data, including a ‘Fair Use’ and an ‘Opt In Unless You Opt Out’ approach to data disclosure.

Policy is around data protection and exclusive rights on data, but what is needed now is regulation around data use providing incentives to share. UK policies enabling a

¹⁷⁵ <https://www.theguardian.com/local-government-network/2013/apr/26/councils-hack-day-geek-squad-problem-solving>

¹⁷⁶ <https://www.theguardian.com/local-government-network/2013/nov/11/big-data-camden-council-digital-revolution>

trusted sharing of personal and business data is essential for new and innovative business models (digital entrepreneurship) to take off in the UK. It is also the only way for individuals to reach the benefits from Big Data, Internet of Things, Artificial Intelligence (AI) and most other digitally enabled disruptive innovations. We need to make the smarter society a reality, and public sector should lead. One estimate in a report - The Value of the Digital Economy - by the consultancy BCG is that the applications created with personal data have the potential to generate as much as €1tn of value in Europe annually by 2020, with a third of the total flowing to private and public organisations and two-thirds accruing to consumers. But for this value to be unlocked, the public and consumers need to feel comfortable about sharing their personal information. They need confidence and trust in the organisations that hold their data, in particular, that the conflicts of interest, privacy and ethical issues will be addressed, and that proper redress is available when there are problems, transgressions or grievances.

By introducing a 'Data Charter' on what can be done with personal and business data, everyone will know how their data is used, which in turn increases trust and creates incentives to allow data to be shared. This Charter would mean a shift from policies around controlling the data itself to how the data is governed. As a first for Europe, the Data Charter should actively send proposals to the European Union to advance into EU Data Protection legislation and harmonisation across borders.

The Data Charter should be used as a reference for AI Ethics Boards in companies to set transparent principles on how data will be governed. It could also become the basis for a Consumer Data Watchdog dealing with data issues around which consumers can unite enforcing trading standards surrounding their data. In this context, we go as far as proposing an international code of ethics that the UK can take leadership in establishing and then promoting at a global level. Data use is clearly central, but the ethics of AI use is a broader subject. There have already been works in this area, such as the Asilomar AI Principles created last year.¹⁷⁷ However, these principles should be further developed to ensure their practicality.

¹⁷⁷ <https://futureoflife.org/ai-principles/>

The Code of Ethics should build on the Data Charter and further set forth the standards for how AI technologies can ensure social impact. They need to be created with the collaboration from policy-makers, corporates (big and small), academics, and the wider public.

Such a Data Charter should also introduce 'fair use' of personal and business data if people are not competing with the owners of the data or harming their ability to monetize it. This 'fair use of data would create a genuinely free space to innovate by supporting entrepreneurship from the data revolution.

There should be equal access to data platforms or shared information systems on which AI data be retrieved in a user-friendly way by the public, so people can know their public record and benefit from knowing information about themselves in a structured way.

Finally, the Data Charter should also adopt an 'opt-in unless you opt-out' approach to personal and business data disclosure. Allowing citizens from birth to be born into a data sharing revolution (in which there is a Data Charter governing the use of data including how business can deploy private data) will empower each citizen. Just as there is no point in being the only one with a telephone or on Facebook, people and companies could only capitalise on the opportunity from personal data when it is shared.

The solution: APPG AI's call for a UK landscape review (see Evidence 1).

In the need for change in data governance, the APPG AI recommends an analysis of the current legislation landscape in regards to data-related issues, including the General Data and Protection Regulation (GDPR) that is to be put in force by March 2018.

3. The Role of Government - What role should the Government take in the development and use of Artificial Intelligence in the United Kingdom? Should Artificial Intelligence be regulated? If so, how? (Q10)

The UK should have an ambitious and trusted 21st Century UK data infrastructure, which supports the growth of the AI based economy to benefit the private and public sector alike (see Evidence 2 and Evidence 4).

Background: The UK's industrial strategy, regional strategies, and infrastructure investment is operating without diagnostic tools or proper context. There are reasons for government analytics to become a lead user on AI.

We often talk about how Artificial Intelligence (AI) changes businesses and consumer relations. However, we talk less about how AI and these new 'market institutions' affect public services, governments (national and regional), and policy makers making decisions on policy, regulation and the budget. It is in the public sector that some of the biggest new opportunities from AI are to be found, but (as reviewed in above section on Data Charter) this requires rethinking new rules, norms and standards in how data are collected and used. If the UK is to lead in this revolution,

But the data revolution with artificial intelligence go beyond public services. It is the foundation of our economic planning. Clearly, a 21st-century government reporting framework on the economy, productivity measurements and the regions, should capture the performance of the current stage of affairs. But the UK data system is technologically outdated, costly to run, and a methodically past. In consequence, the numbers wrong or useless.

As a result, the government cannot properly plan its budget, infrastructure investment, tax levels, and public expenditure for research, education, skills and social issues. It also has difficulty in deciding the sectors and technologies around which to develop support strategies. Business leaders cannot even themselves set sound strategies for their investment and performance efficiency challenges.

Firstly, government data collection and measurements do not capture knowledge-based services, new forms of manufacturing, and the digital economy including the effect of new forms of work, automation, smart devices, robotics and artificial intelligence. The conceptual, theoretical and measurement frameworks developed for a physical paradigm and past industrial revolution need re-addressing.

For example, productivity measures used by national income accounting focus on quantities produced and physical measures such as machinery, buildings and hours worked. The dimensions of quality, sustainability and service generated by intangibles are not captured even though they are vital to successful company investment and government policy alike. Productivity measures are outdated, fitting better to the post-war industrial economy than today's knowledge-based digital economy.

For instance, today, energy services are meant to improve sustainability but productivity is still measured by how much energy is physically sold. So while energy providers invest in high-tech, supplier

networks and manu-services that help consumers save energy, productivity is still measured by the quantity of energy delivered. Energy firms want to help consumers economise on their bills, but the more successful they are, the slower the growth in sales of electricity and gas. Consequently, the productivity growth as conventionally measured would be slower. Similar for financial services: Productivity measures should not be grounded in the number or size of transactions (loans and cash accounts), but how well the banks manage people's finances or that of the economy. Productivity, in short, needs rethinking.

Energy, health, transport, finance and retail are five major sectors where consumers are expecting improved quality and sustainability as opposed to more quantity. Most contemporary value added to work is the deployment of intellectual capital in production, services and manu-services: here people do not produce more 'stuff', but increase its quality.

Secondly, the design of data collection structures is not fit for purpose, but segmented, analogue, and hold gaps. For example, there is a lack of data input to the Office of National Statistics, Companies House, Treasury, and Bank of England. Data supplied by large multinationals are better captured, but data collected from SMEs and Public Sector organisations are missing or incomplete. The same can be said for data collections from EU data (CIS) and the OECD. Problems are especially around the missing 'innovation systems and intangible asset' data.

Also, data are not collected for a specific purpose – as for example to develop our Industry, start-up or talent systems.

The solution: Artificial Intelligence, with associate analytics and diagnostic tools, should be used as a tool to inform the Industrial strategy and the Budget

A strong data infrastructure means integration of public and private data collection sources on one platform (information system), an upgraded focus on innovation and intangible asset data, and direct link with stakeholder use and purpose.

Economic analytics models should be updated, given that the current ones are modelled on the features of a past economy not taking advantages of the internet and Artificial Intelligence.

There are lessons to be learned from US, India and Europe, but China's vision of economic development from economic data is inspiring.

Transforming our regions and our supply chains to become innovation hubs like Silicon Valley, Boston or Bangalore is a major aspiration for

the United Kingdom. There are global exemplars of what works. Whereas Silicon Valley and Boston developed with close links around world class Universities, Bangalore developed with close global supplier links to Silicon Valley until it became a thriving hub in its own right. Eindhoven, located in a much smaller provincial part of Europe, took a different route with Philips Electronics (a big corporate) as the hub – but with a good-enough local university and looking to outsource IP and technology to an innovative supplier network. Philips Electronics crowded in expertise from world class academics – often created a link to the local university - and opened space for entrepreneurs to co-create with them locally. They invested in new buildings and converted outdated factory space 'not fit for purpose'. All the approaches created opportunities for the local regions to upgrade. However British regions have few comparable assets, nor have our own efforts so far shown much success.

China has taken a different, more systemic approach – what it characterises as an 'Opening up of the system' approach for regional and economic development, transforming regions and cities with high tech clusters, industrial parks, and taking millions of people out of poverty. The method included development from economic data and 'achievements from system construction' (as opposed to classic macro-economics). CEO of Big Innovation Centre has visited six Chinese regions and believes there are lessons to be learned.

Using the lessons learned from these international models, Big Innovation Centre has piloted diagnostics tools capabilities using artificial intelligence for an online-real-time assessment of the skills-base and the innovation capabilities of the UK regions, across an agreed set of industrial and entrepreneurial segments which supply our business, trade and job base. We also investigate the capabilities of the education and talent system which provide the skills-base for the future. We address the capacities of our transport in travel to work places and infrastructure system as well as highlight areas of deprivation with respect to health, crime, access to opportunity and culture. Do contact us at info@biginnovationcentre.com if you want to know more about this initiative which can lift the capabilities of government agents (national and regional), local businesses, universities, property developers, and investors with enhanced decision- making capabilities.

There is much debate on whether AI should be regulated or not.

The APPG AI community seems to be split on the matter: some calling for changes in legislation and others pushing for the use of soft-structures to address AI-related issues.

Big Innovation Centre advises that before any new regulations are put forward, Government needs to gather fact-based evidence which

exhaustively analyses the current impacts and anticipates for future repercussions (short-term and long-term). In this fast-paced environment, the evidence gathering must be quick and practical. The European Parliament, in June 2016, published an overview of EU laws and rules that will be affected by developments in the fields (AI, robotics, cyber-physical systems), identifying 39 EU regulations, directives, declarations and communicates that may need to be revised or adapted.¹⁷⁸ We propose that the UK adopts a similar methodology to further understand the landscape before deciding which changes in regulations are necessary (if any).

4. Public Perception - Should efforts be made to improve the public's understanding of, and engagement with, Artificial Intelligence? If so, how? (Q5)

Yes – According to the AI experts from government, business, and academia in the APPG AI community, improving the public's understanding of what AI is – and what it is not – is crucial at this point in history. AI has already impacted our nation economically and socially, and its effects are anticipated to skyrocket in the upcoming years. In fact, according to a June 2017 report by PwC, UK GDP will be 10.3% higher in 2030 as a result of AI.¹⁷⁹ However, although AI is likely to be one of the most transformative and disruptive forces of the decade, the public lacks a basic understanding of what it is. Based on the April 2017 survey by the Royal Society, only 9% of the respondents recognised the term 'Machine Learning'.¹⁸⁰

The public's understanding of AI can be improved through various channels. We propose (see Evidence 1):

Education: The UK should introduce the term AI to children from a young age, explaining to them what it is, its opportunities, and its risks. The topic should be included in school curriculums using appropriate language that is inclusive, accessible and accurate. Most importantly, children from an early age should start building STEM (Science, Technology, Engineering, and Mathematics) skills necessary to compete in the modern world of AI technology. Private companies such as NVIDIA are already creating workshops targeted for students to get a basic understanding of how neural networks work. Similar initiatives should be promoted

¹⁷⁸ <http://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence>

¹⁷⁹ <https://www.pwc.co.uk/services/economics-policy/insights/the-impact-of-artificial-intelligence-on-the-uk-economy.html>

¹⁸⁰ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-IPSOS-Mori-summary-report.pdf>

in order for upcoming generations to have the information needed to understand AI and, consequently, make well-informed decisions on its implications. Educational reforms and campaigns should also be used as mechanisms to inform older generations that are already in the workforce or about to enter it. AI should also be embedded with formal and informal curriculums for higher-education and lifelong learning programmes.

Media: The media is another powerful tool to inform and educate the public on AI-related issues. According to the speakers at the APPG AI meetings, most media stories have focused so far on the negative implications of AI. People tend to think of AI in terms of science-fiction movies that tell a story of a robot killing mankind or a news article foreshadowing huge amounts of job losses. Media should portray both sides of the conversation, also shedding light on the opportunities and benefits AI technologies will undoubtedly bring to society.

5. Summary

In order to become a global leader in the field, the UK Government must act quickly. We have the power of shaping the future with AI, but first, we need to make sure we propose fact-based, pragmatic solutions. To reap the full benefits of Artificial Intelligence we propose:

- A Data Charter, on what can be used by personal and business data, including a 'Fair Use' and an 'Opt In Unless You Opt Out' approach to data disclosure.
- An international code of ethics, setting guide lines for corporate AI Ethics boards and a consumer AI Watch Dog.
- A review of the current legislative AI landscape in the UK, mirroring work already done with respect to the EU.
- Government to become the lead user of AI, especially with respect to upgrading public services.
- To replace and modernize our economic tools with AI diagnostics, build to inform the Industrial Strategy and the Budget.
- To modernise UK's data collection and reporting infrastructure be fit for purpose in the new 21st century AI based era.
- To introduce AI into both formal (schools) and informal (training) curriculums, and inform the public of AI opportunities through media channels.

Big Innovation Centre

6 September 2017

Bikal – Written evidence (AIC0052)

DATE: 3rd September 2017

TITLE: Submission for Call for Evidence to the Select Committee on Artificial Intelligence

NAME: Mr. Raj Sandhu

CEO of Bikal, a company that supplies High Performance Computing for on-premise or private cloud installation. I am commercializing the technology generated through academic research by co-designing it with Artificial Intelligence capability.

I work with domain experts, retired experienced professionals, that identify industry problems where parallel processing of data in real-time can benefit the end user. This is specializing in analytics of multiple data sources over a period, and creating automated features to aid the operator.

Our partnerships with universities and organisations mean that we provide an agile testing procedure for research to be transferred into the real world.

The pace of technological change.

1. Tech evangelists are stating that PhDs do not now need four years to complete, due to the power of computer modelling. Research theories can be assessed in days rather than weeks. The consensus is that 6 months is sufficient. Retailers are progressing to larger sets of demographics to segment their customer base to make better predictions, however, they are some distance away from the segmentation of one, but it is possible to do this now. I would suggest that regulation, such as GDPR, will empower a greater relationship with companies and their consumers (where it be retailers or councils and their citizens). Incentives for consumers will drive greater and broader data sharing if the processing company can make accurate predictions for that individual, and provide adequate incentives. For instance, the cost of finding someone's home insurance renewal date is estimated at £90. An obvious incentive is to pay £50 for that data. But if an insurance company shared that data with a building company, and a furniture company then the consumer may get further benefits in savings on home improvements if they were targeted at the right time. Currently, researching a product on Amazon means that you will be continually be targeted on that product even if you purchase it elsewhere or decide not to buy it. This will change as compute and storage costs come down further, coupled with changes in attitude of personal data being processed.

The Role of Government.

1. Information Sharing.

Companies process data, largely, in siloed environments and tend not to share data. Purchasing of retail data, with end user consent, by marketing agencies is processed for targeting individuals based on location and or products and services. This is typically mis-timed and still inefficient. Government has an advantage in that it has historical and varied (from multiple agencies) data sets that can be exploited for their own benefit, test algorithms for regulation, and commercialise.

2. Testing AI with government data sets.

Our work with tech transfer from universities has given us access to algorithms that have been researched using multiple data sets, which in the commercial world is very hard and lengthy to negotiate. The tendency for companies is to state that the data protection act stops them from testing, which is not the law.

Using the appropriate legal sections, we have opened discussions with social services to coordinate information sharing between their associated agencies to be able to use our algorithms for pattern detection and prediction of outcomes. Using the historical data from all the agencies associated with social services we will be able to apply Artificial Intelligence so create a self-learning system that can look for patterns or negative and positive outcomes. These outcomes rely on input from the operators, especially the ones on the ground This will enhance the weekly meeting between agencies on making assessments on their cases, without having to rely on individual knowledge of a case, and more significantly, relying on scraps of information that on their own do not appear to be important or critical. Private industry cannot operate on this level as easily as government. For instance, companies in the retail sector would need to research which industries and to what extent an individual is influenced to purchase their product. In the public sector, a person's employability could be determined by their age, health and other social factors which the council can analyse through data they hold.

3. Education.

Companies and local government should have some form of incentive, perhaps as part of a R&D tax credit, to introduce AI to their employees. Online courses are sufficient to provide an insight as to how AI is developing, and how it can create new opportunities by automating and improving existing repetitive processes. The workforce of today will be working for longer, and competition from generation Z and millennials will most likely be based on creative, and data science skills. Our engagement with a large energy supplier on AI has been based on a change in strategy, where it has essentially become a data company that supplies energy. It is cannibalising its own revenue by using AI to predict power usage so that the residential and business consumer has a lower bill. They have identified, or perhaps accepted, that if they do not do this then a

tech disrupter will take it from them. Having the strategy for increasing margins by reducing revenue is a brave step and embraces the Industrial Revolution 4.0. These savings were made by training the current operators on the benefits, not features, of data science to their industry. Their supply of data scientists and computer scientists is through collaboration with a local university, where the head of data science at the energy supplier helped design some of the courses for AI that would be used at their company. They provide one-year placements for students and have ad-hoc hackathons for business challenges.

2 September 2017

Dr Richard Billingsley – Written evidence (AIC0201)

The state of artificial intelligence

1. Artificial Intelligence (AI) took a major step forward in 2006 with Geoffrey Hinton's development of Deep Learning. It is now growing rapidly and commercially being used by most major information processing technology companies.
2. AI can now recognize a person's face, a type of fruit, or transcribe what you were saying. It can translate text from one language to another, summarise a story and answer questions about a picture. It can detect abnormalities in medical images that previously required a specialist. It can spot unusual activity in a CCTV feed and extract sentiment from twitter feeds. It can predict where you would like to eat, what you might like to buy, your mood and emotion from your remarks or appearance.
3. AI can generate realistic images, win at board and video games and teach a biped to walk and run from scratch. It can operate machinery and drive a car. It can make predictions for trading in financial markets and for making longer term investments.
4. AI is consistent, can work tirelessly without fatigue and produce reproducible results after digesting unimaginable quantities of information. It can combine information from complete strangers or acquaintances on diverse topics to learn insights about an individual that any person could never otherwise discover. Already, AI replaces people's jobs. Every item found and bought online was a win for AI and a loss for a sales person.

The next 5 years

5. AI is the junction between data and algorithms. Data has never been collected more rapidly and the pace of algorithmic development is at breakneck speed. In five years we can expect most AI development to still remain mostly invisible, hidden in data-centres powering search engines, phone apps and AI agents, like Alexa and Cortana. However, they will be drawing greater insights with ever improved clarity and understanding, probing and learning every facet that can be guessed about their users.
6. Corporations will use more AI for automatic question answering, image inspection and analysis. What's wrong with my fridge? Take a photo and get an instant answer. AI will be handling more consumer to business phone calls and making more automatic replies to emails.
7. CCTV footage will be analysed automatically. Immediate alerts to the police, fire or ambulance of problematic activity will be sent whenever an algorithm spots something amiss. More bureaucratic processes will be

automated through AI. Application forms for immigration, permits and licenses will be processed by AI algorithms with only edge-cases examined by human eye.

8. There may be visible signs of AI, like robots in shopping malls showing you to the shelf, but these will still probably be for marketing purposes to wow and lure customers in a largely scripted and tame fashion.
9. But the speed of AI progress will be ominous. Most video games will be conquered by AI algorithms. Stories, scripts and music will be generated by software with quirky results. Search engines will become more prescient as new algorithms are combined with old data, never deleted, leaving individuals to give up on privacy concerns as a lost cause.

10 years

10. The mind still learns more rapidly, more nimbly and more effectively than any deep learning program. Giraffes can walk moments after birth, and animals recognize danger even from the first encounter. These biological successes show that AI algorithms still have breakthrough potential which intense research effort will most likely uncover within 10 years.
11. This would lead to robots in the home cleaning, doing laundry and babysitting; Robots in the field gardening and picking fruit. Mindless, boring and repetitive tasks being taken over by machines that watch and learn and send their insights to be processed, shared and commercialised.
12. Other outward signs of AI like self directing drones navigating point to point, self driving cars and trucks and machinery can be policed more easily, and may be delayed by legislative processes. But information cannot easily be policed and travels between borders. Uber drivers might be pre-warned of likely customers using demand prediction and mobile phone location data.
13. AI may be writing music and poetry of a commercial quality. Algorithms to write legal briefs and quality legal arguments may be used in court cases. Photorealistic movies may be made entirely by AI. AI will be the dominant factor in production line automation, and the days of the low-cost-labour advantage will be ending. International economic flows will consequently be shaken and AI will be the dominant factor in espionage.

20 years

14. Within 20 years, everyone will consult their phone's AI to make even mundane commercial decisions. Corporate strategy will be guided by AI and the independence of AI to outside influence will be crucial.
15. AI will be the dominant factor in the military, with drones and robots rivalling nuclear capability in importance. News and opinion will be shaped

by AI by predicting public responses to statements, so guiding corporations and politicians on what to say and when. We will be as dependent on AI as we are on vehicles. You can live without it, but can you get anywhere or compete?

What factors, technical or societal, will accelerate or hinder this development?

16. AI progress will accelerate due to the large network effects caused by demand side increasing returns to scale that leads to monopolistic outcomes and makes the rush to be first important.

17. The difficulty to regulate this borderless industry and the ease of bypassing patents through black boxed non-disclosure allows rapid progress in research. Likewise, the ability to bypass privacy laws through one-click agreements, and the potential to sell a venture to a large buyer thereby selling the amassed private information makes AI development and data harvesting even more valuable. Little will slow the race to be first.

Is the current level of excitement warranted?

18. AI has the potential to be the most impactful development of history. Like the printing press, telegraph, telephone and internet, it deals with information. Like the bicycle, car, microwave oven, toaster and fridge, it makes things easier for people. Like the combine harvester it has the potential to replace many people all at once.

19. It has the potential to impact industries, international economies, politics, unions, society and the workforce. It will change the way we do things more than the computer and internet ever did. The excitement has only just begun.

Benefits of AI

20. AI will allow mentally and physically tiresome tasks to be performed with ease. This will help increase the GDP above the rate of population growth and inflation as more can be done with less. As inflation is dependent on the poorest 5% through NAIRU, its impact on inflation and asset prices will be determined by the accompanying policy shifts.

21. Given the choice, most people would much prefer to take it easy while a machine performs their work than do it themselves, provided they still get paid the same. Likewise, a country equipped with AI will outperform one without. Consequently, AI is a boon to every economy and individual provided the economic and regulatory framework adjust quickly to prevent it becoming another tool to accelerate income disparity.

22. The undoubted benefits of AI helping humanity should not be deferred simply because of the economic hardships that will ensue. Instead, the

economy must be rapidly realigned to ensure everyone benefits from its arrival. We should separate the three goals of controlling AI, improving productivity and preventing inequality.

23. Indeed, the high debt state of the economy might desperately need the improved productivity that AI will bring to help make debts more manageable. But it is easy to see how present day low cost manufacturers and labour forces will lose their competitive advantage. This will cause global changes in trade flows, job prospects and skill requirements.
24. The goal is not to stop stealing jobs (a Luddite approach), but to stop stealing the pay. Most would be happier if they could work less for the same relative pay. A greater disconnect between work and pay will remove the pressure of work. When robots make labour less scarce, the benefits for an economy to motivate labour is reduced so the economy should be re-engineered to favour greater national and global satisfaction.
25. Asset price inflation has already disassociated work from pay, with millions of home owners earning more from their house than their work. Income tax is only a relatively new invention.
26. Solutions could include parental payments for the work of bringing up children, a guaranteed living support for everyone, grants for the arts, and higher pay for service work like counselling and aged care.
27. These benefits must be largely paid for by the proceeds of AI so mechanisms must be in place to ensure the monetization of personal data and publicly researched algorithms is not commercially benefiting very few private organizations, but shared nationally and globally. If the inventor wishes to stand on the shoulders of giants, perhaps he should also pay for the view.
28. Inventions are already controlled by patent licenses, an economic construction created entirely through law. Likewise, new economic constructions may be needed, for example, a personal information tax for holding or processing personal information, or a license fee robots must pay to financially insure their operation.

Impact on society

29. AI will impact everyday life in small, almost invisible ways. Some products will become cheaper, easier to use, or more fun. But many of the changes will be designed to extract more information about you. Video games may have cameras that can see where you live. Gadgets that record you, your heart rate, what you see, who you talk with will proliferate. Self driving cars may send destination information back to servers. Train tickets may record who bought them so personalised routes can be mapped. Complex

economic decisions will particularly be analysed like mortgages, education and car choices.

30. Consumers in general will benefit from AI with better smarter services. Employees who touch or meet their client will generally benefit from or be immune to AI's arrival, but cubicle and outdoor workers will see competition from AI devices. Licensed occupations like doctors or plumbers will be far safer than unlicensed ones like factory workers.
31. School education may remain largely unchanged with the addition of an AI class, but should ideally focus much more on soft skills like teamwork, leadership and social interactions to drag children away from their screens. Human interaction work will flourish, like psychology, counselling, kindergarten teacher, art dealer, estate agent, salesman. However, human individuals may turn to their more entertaining gadgets while real-world social interactions suffer.
32. Society must focus on being more cohesive. The individualizing nature of customized AI encourages segregation of thought and ideas. Social customs will need to address this by creating more mixing events, market days, opportunities for different people to meet and exchange ideas.

Data monopolies

33. AI has large network effects. Products with better AI perform better so become more popular. More popular products collect more data so learn better AI. This leads to powerful monopolistic effects with few large private companies amassing giant data stores far greater than many governments.
34. For example, combining email and location data with insurance claim data enables powerful analysis of which activities, places and foods improve or impair the population health. This kind of information would help people plan activities and diets or isolate causes of infection outbreaks. Enabling thousands of independent researchers to find these insights and build businesses with products creates a public good, while keeping the data and outcomes private and concentrated in a few businesses simply distorts existing markets.
35. The closed nature of AI development is confirmed by the proportion of leading AI experts who have left academia to join a handful of multinational data titans, often citing reasons that only such places have worthwhile data for useful products, algorithms and analysis to be developed.

Who gains the most from AI and data

36. AI most benefits those with the largest stockpiles of data which leads to better commercial decision making. Companies with the most to gain will consequently advocate for open source development, peer review public disclosure of new algorithms and sharing for the good of everyone.

37. The strategy of the smaller data-light start-up will be to develop in secrecy using their algorithm as the barrier to competitor entry, rather than their data stockpile. With AI's potential for military uses, patents might be circumvented, so inventors may adopt a gold-rush mentality of distrust and secrecy with advances occurring unknown to all but a few.

Who gains the least

38. On the losing side, individuals on whom data is collected lose out. They get the least valuable insurance when insurance companies better predict who will fall sick. They get less favourable financial products when banks better know who is a safe bet. Fewer commercial opportunities when they are automatically mined and sold to the highest bidder.

39. Companies in competitive businesses also lose out when AI insights prove valuable, but can only be bought at monopolistic prices. This cuts in to their profit margin making the insights invaluable, particularly if the AI providers also control the advertising channels.

40. Researchers and small companies will lose out unless they have access to the large privately held datasets. This will give them little choice but to join or partner with the large data holders, and then with little bargaining power.

How disparities can be mitigated

41. To avoid large disparities, the many benefits of AI must be shared as freely possible. This requires that no-one has a monopoly or near monopoly on being able to analyse the data, choosing which parts to share and which to keep private. One must question how it can be good for one private organization to have exclusive access to this data, and not two or three.

42. Consequently, it is important to spread out the data so organizations must collaborate in order to share these diverse data. If single organizations could not collect both location data, search data and robot sourced data or could not both collect and analyse data, this would create more public good in the AI data marketplace.

How data can be managed so it contributes to the public good

43. Previous monopolies like telephones have been broken up into separate industries, one making handsets, another providing connectivity, and

another providing customer servicing. Big data could also eventually go the same way.

44. Data organizations hosting data above, say, a thousand terra-bytes of personal information could be regulated in how they handle direct personal data (the original data) and any derived data (that could be learned from the direct data). Regulations might prevent them from mingling different user's data or distributing any analysis or derived data except through regulated contracts to third party organizations. In this way, the storage and processing of data could be separated reducing monopolistic practices.
45. These regulated contracts might prevent single buyer agreements, to provide competition in the data analysis marketplace. In this way, large and small companies and government agencies could still get equal access to suitably redacted personal information that only a few large entities presently retain and on-sell.

The ethics of AI

46. People may not be aware of how much their private data is collected. They may assume that once logged out or with browsing history turned off, your searches are no longer being recorded and processed. Passively accessing websites can result in consolidated browsing histories being known and analysed thousands of miles away.
47. Phones are recording your location in real time, and can tell who you travel with, where you go, and how long you stay there. If you stay longer in the doctor's office than usual, some server knows about it. If you go for an MRI, visit the casino or invite your friends for the evening, all these details can be deduced from phone location data and be legally sold to interested buyers.
48. You may think your data in the cloud is safe and private, but the hosting company can be bought out by a data aggregator so the once private data is analysed for commercial gain.
49. Robots will provide a far greater invasion of privacy. Like wearable cameras, they provide the opportunity for live data aggregation of what you do, wherever you do it. If you are on the beach, a passing robot may capture video of you there. If you are cooking in the kitchen, the helpful kitchen assistant is sharing your every move with servers off shore. Big brother will definitely be watching, but will be owned and operating as a commercial entity.

Consent

50. Privacy is completely lost if using an essential product requires signing away all your protections. If every water tap said 'by using this water you agree to give up all rights to privacy' would such agreement be fair? Can

we compete and survive without our phones, email, social media and search engines? AI is becoming a necessary part of life, so agreeing to terms and conditions is becoming less optional.

51. Who knows what terms you have agreed to in order to use your phone? If you don't know them, did you really agree? Without a signature, can anyone prove that your three-year-old child wasn't the one to click 'I agree'?

52. When watching movies, you consent to be shocked or bothered after noticing the adult theme certification rating. Likewise, large data services could be rated:

- o We [do/do not] sell embarrassing analysis of your email, posts, queries or robot data
- o We read/scan your content and keep it indefinitely for future sale
- o We own your data and may sell it with other assets if we are bought out

53. A standardized online agreement could require users check to confirm they understand the above points in clear, common language to ensure they are not duped but realise the implications of the broad terms of 'we may use your data to improve our products and services'

Diversity

54. AI is based on learning from patterns. This promotes the perpetuation of stereotypes. If banks think white males are more successful in business, then AI equipped banks may compete harder to offer them cheaper capital for their business, even if white males improved measured success is entirely because they can access cheaper capital.

55. As AI learns from mining public data, it learns every public prejudice. AI can be used to avoid culpability for racial profiling. Instead of filtering against a minority, it can filter against where that minority predominately lives or goes, or what they do or the word phrases they use.

Democracy

56. Each child tacitly learns to accept a framework of laws crafted throughout a turbulent history that are essential for civilized life. But if technological changes lead the existing framework to yield inequitable outcomes, the laws must adapt.

57. Societies that are able to prosper without the success of a large subsection of their people can better afford to marginalise and disenfranchise that subsection. In this way, oil rich nations can better afford a less engaging democracy when they do not need the hard work of everyone to succeed, while grass-root service based countries must look after everyone to thrive.

58.As AI replaces the need and value of labour, asset prices will rise increasing inequality. Just as imprisonment is not justified by jails but by the equitable administration of justice, so too the advent of AI and situational awareness might enable but not justify the control of the population unhappy with rising inequality. It is doubtful even the Boston Tea Party would have gone so well with live streaming CCTV cameras and AI facial recognition, but many would argue that its success was a long term public good.

59.As AI firms know the hot-button issues of large sections of the populace, it will create suspicions they could influence these issues to invisibly sway elections. It will require determination to overcome the challenges to ensure a more democratic national debate and more aspects of technology will be needed to provide frameworks and strategies to protect the virtues of a free and fair society.

Equality

60.AI has the potential to greatly help people, but scenarios can be imagined where this is not the case. Car route navigators could learn congestion management to prevent all cars taking the same route. It could then give preference for the faster routes to paying users, drivers of Mercedes, or those who wrote favourable tweets. It might route you past the store that paid the most. Petrol prices could change as your license plate identifies you and your purchasing power or desperation for fuel. Shopping prices might be adjusted according to your purchase history. Your insurance company could learn how long you visited the doctor and reassess your premium.

61.AI may also remove accountability. Unfair decisions can be taken in microseconds which are arduous to appeal. For example, AI used in traffic policing may determine you crossed a red light, but these should be easy and un-bureaucratic to appeal.

Black boxing

62.Much of AI uses unrecognizable latent internal structures so is inherently black boxed. Avoiding culpability for discrimination and data leaks is a major concern. Suppose a black-box trading algorithm that makes trades based on an expensive or exclusive statistical data feed that includes information from a comparison shopping website that can be used to detect whenever a company has increased their advertising budget. Did insider-trading take place and if so who was guilty and could it be policed?

63.The Brecon Beacons shows how a single bit of information can prove very useful. A venture capitalist planning to invest in a start-up would likewise

benefit from knowing if the founder has distracting marital difficulties, pending lawsuits, or spends too long at the golf club. All this could be extracted from his email and phone location data, never seen by human eyes but refined by AI to a single invest/don't-invest SMS.

How to regulate AI

64. Artificial intelligence itself is a collection of mathematical calculations and algorithms that can be run anywhere from a laptop to a server farm. As such, they know no borders and are impossible to police, so it would be impossible and foolhardy to try to regulate their development and use alone.
65. The scope for any regulation will be limited by international competition and the borderless flow of information. If AI cannot be performed here, it can easily be performed overseas, bringing the cost savings and economic benefits to foreigners. To overcome this, a coordinated leadership approach with other likeminded countries will be needed to create the economic scale to guide AI regulation.
66. There are three stages of a problem. The first are those so small you should do nothing. The second are those so serious you should do something. The third are those too large to be addressed so you should do nothing. It is prudent to address the control and development of AI while in stage one, before it slips quickly through stage two to stage three.
67. AI will centralize and take control of vast amounts of personal information. Care must be taken to ensure that countries do not lose their sovereignty by knowing less about their citizenry than foreign multinationals.

Privacy

68. Email service providers could be publicly rated according to the privacy they offer. The post-office or government could provide everyone really private email accounts with Government, health, school and tax business only conducted by email with accounts above a certain privacy grade. Private data can be regulated requiring it not be sent offshore to more lax jurisdictions.
69. The protection of private and personal information may already be the subject of privacy laws. But derived parametric data, the result of processing, aggregating or generating parameters through learning from private data and other derived data, might not be. Mathematical approaches to measuring how diluted information has become after processing exist and thresholds could be established for protecting data above certain privacy concentrations. Using these would allow for regulations protecting the output of any process that supplies semantic content of private data.

Physical

70. Robots may take a very different form. You might bring your robot's mind on a thumb drive or phone and connect it into the shopping trolley for shopping, and to the car for driving home. So another key issue may be tax jurisdiction. When a man works in a field, it is clear where the work is performed as body and mind accompany one another. But a robot may have reflex actions performed locally while the main mental challenge occurs on a server offshore. The robot may be being instructed by other computers or people in separate countries again. In which jurisdiction then is the robot living, taxable and governed?
71. Likewise, robots may provide telepresence for offshore workers, allowing them to bypass immigration queues and long haul flights to work here from overseas through virtual reality. The government's ability to control workers' rights and minimum wages would quickly evaporate when borders become porous to virtual labour.
72. One possible regulatory route would be to license all robots above a certain weight or power consumption like motorbikes. Unlicensed robots would be uninsurable while licensed ones could have some legal protection from accidents and third party privacy infringements. This could state the legal responsibility for the robot and its jurisdictional issues.

The role of government

73. Government must set laws and regulations to ensure the most competitive adoption of AI leads to more public good. AI offers enormous benefits which must be shared equitably. AI and robots will have a significant role in society, the economy and development of the nation. Investments should be made to:
- o Stimulate the local development of domestic research accessible databases
 - o Fund access to data-centres for research and development
 - o Promote AI components in government tenders to accelerate local development
 - o Address issues of cohesion in society and technologically induced inequality

7 September 2017

BioCentre – Written evidence (AIC0169)

Introduction

- BioCentre is a think tank based in London looking at the ethical, social and political implications presented by new emerging technologies. The mission of the centre is to be recognised as the place which is 'hosting the conversation' concerning the major implications posed by emerging technologies as they impact upon the future of humanity. We are concerned as much about the conversation and *how* the questions are framed as we are about the answers. In so doing by fostering a cross-disciplinary knowledge network, BioCentre seeks to clarify and frame the key questions, providing informed opinion and advice on these advances.
- Engagement in the conversation surrounding artificial intelligence (AI) and robotics has included a recent series of horizon scanning public symposia and consultations on the use of AI and robotics in caring for the older person. BioCentre has also held a public discussion on robotics and the future of work with University College London's (UCL) Science, Medicine and Society Network and an afternoon consultation with maritime trade unions.
- BioCentre's Executive Chairman, Professor Nigel Cameron, has also written on the subject in his recently published book, ***Will Robots Take Your Job?: A Plea for Consensus*** (Wiley 2017). Key material from this book is quoted from and referred to in this submission.

Definitions

- a) The responses offered in this submission are based on the following definitions and understanding of artificial intelligence (AI) and robotics.
- b) Work in the field of AI is concerned with creating a computer 'mind' that thinks like a human. This challenge has been the focus of many scientists and technologists for decades with varying degrees of success.
- c) Originating from the work of a science fiction writer in 1920, the word robot comes from *robotá*, a Czech term for servitude¹⁸¹. A robot is a machine that is capable of carrying out a series of actions automatically, often work that was previously undertaken by a human.
- d) Combining developments in the field of AI with robotics makes for one of the most exciting areas of robotics as attempts are made to build 'intelligent machines' or robotic devices powered by 'machine intelligence' that don't look like Robots, or may not have physical form at all.

1. The pace of technological change

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

¹⁸¹ Carr, N. 2015. *The Glass Cage: Who needs humans anyway?* Vintage: London. p225

- 1.1. The key reason why these advances are beginning to be discussed and posing profound questions about the future is that uncertain things are possible. In many respects technology has an agenda all of its own. Driven with great dynamism its influence drives decisions, conversations and tends to drive outcomes unless it is engaged with effectively.
- 1.2. Of particular relevance to the AI and robotics conversation is the Internet of things (IoT). It is estimated that there will be nearly 20.8 billion devices on the IoT by 2020. This network of interconnected devices that will collect and exchange data will help to drive forward the idea of the 'smart home'. It will therefore become the most natural thing in the world to have robotic assisted devices as part of this panoply of interconnectedness.
- 1.3 Moore's Law is driving this digital 'explosion' or 'revolution'. In simplified form this law states that computer processing power doubles every two years. In other words, anything driven by computer technology – anything that can be digitized – is advancing at an exponential pace. While there is evidence that Moore's Law may at last be slowing down (Technology Quarter, 2016), the prospects for quantum and biological computing suggest huge changes ahead. If we consider this rate of progress as a graph we are traveling up, looking back and charting our progress we can see how steep the curve has risen to bring us to where we are now. As we look forward, we perceive the line as much smoother. We put a kink in the curve, and flatten it so that the future keeps going up, but more gradually. We simply don't imagine that the dramatic pace of change that got us here can continue – let alone speed up. We've established a "new normal" as the baseline for tomorrow (Cameron 2017:54). This flattening of the line is actually the very opposite of what we need to be doing in order to prepare for the future. The curve is only going to get steeper. While there are aspects of our lives that have been largely unaffected by the digital revolution (cooking, sports, the weather), anything driven by computer power will be driven up that curve. Try sitting down and making a list of all the things you use your "phone" for, and the dozens of processes and activities that used to be needed to accomplish them. The changes ahead will be faster and greater, and we're in denial if we think otherwise.

The technological versus human imperative

- 1.4 The imperative to pursue technology at whatever cost should not trump the human imperative. As members of homo sapiens, we are still relatively young in terms of biology, as well as geology and cosmology. Talk of the "post-human" and "transhumanism," therefore seems rather premature and should be regarded as a distraction to the more serious questions at stake.
- 1.5 As we consider the continuing evolution of homo sapiens there is a place for non-naïve optimism as to what new advances in technology can afford

us. At one end of the spectrum there is naïve optimism and the belief that technology will solve all our problems; what could be termed the technological imperative. At the other end we have what we may call doomsday futurism, in all its varieties, with a focus on the likely impact of existential threats.

- 1.6 However real these challenges are, we need to ensure we are not simply naïve about a context driven by need; need to contain cost and need to provide care. Rather than naively thinking technology alone has the solution, we need to appreciate the role of humanity with technology.
- 1.7 The greater focus must be on striving for the optimism on the far side of the naïve - and the raft of challenges and risks we soberly see ahead. Non-naïve optimism should become an essential methodological principle if we are to engage effectively in discussions concerning the future, not least those pertaining to AI and robotics.
- 1.8 This sober realism recognizes the fact that technology can bring with it great transformation and improvement to the ordering of human affairs. Yet at the same time it has not solved all our problems: the struggle for human rights, freedom, and humanitarianism continues.
- 1.9 We are a young species, with new tools and as such we begin to cut straight to a core question that underpins our anthropology: **what does it mean to be human?** We need to remember that we are not only Homo sapiens or 'wise man' but also Homo faber - 'working man'. The whole idea of technology, from the most primitive tool to the latest silicon chip, is the story of us making things that enable us to do more than we could do without them. It is perhaps no surprise to find us in a place where robots and AI powered machines are being made to copy what we do it and do it more efficiently. Nevertheless, we need to identify and distinguish ways to clearly state that our intention in using AI robotic devices is to enhance the human experience.

2. Impact on society

How can the general public best be prepared for more widespread use of artificial intelligence?

Labour and employment

- 2.1 We need to be wise that we do not end up becoming slaves to the very technology that is supposed to be serving us. It was this recognition that drove C.S. Lewis, in his prophetic essay of 1943, "The Abolition of Man," to argue that while technology is said to extend the power of the human race, "what we call Man's power over Nature turns out to be a power exercised by some men over other men with Nature as its instrument". By taking to ourselves the power to determine who we shall be, we turn ourselves into creatures of our own design, artifacts of our own manufacture.
- 2.2 In the specific case of labour and employment, we need to question and engage with the debate as to whether AI and robotics can help us work more efficiently and productively, giving us more time to do what only humans can

do. The idea of a 'work free' world might at first sound appealing but work can be a means by which we learn, develop skills and overcome challenges. Robots and AIs could support this but can they and should they be allowed to supplant it?

- 2.3 A report by Pew Research explored the views of some 2,000 experts on AI, robotics and economics, concerning the role of automation between today and 2025¹⁸². There was an almost perfect split in opinion: 52% predicting an optimistic path of the future, in contrast to 48% who expressed concern and worry about the future.
- 2.4 Some argue that robots will create more jobs than they will take over. Whilst others worry that their arrival in the workplace will lead to a break down in society. To give just one example, as Google's work on self-driving cars continues to develop, it is not difficult to see that the next big thing could be automated driving, threatening the work of many taxi drivers, lorry drivers and others employed in transportation.
- 2.5 This in turn speaks to questions of inequalities. Because the fundamental inequalities of wealth and income that shape and at their extremes disfigure our human society will before many years pass come to fresh focus with the mounting crisis over the loss of jobs. The opportunity will be open for us to come up with, as it were, non-socialist models of re-distribution, in light of (a) the basic social and market need for most people to have income, and (b) the baseline shift into an economy in which labor participation in wealth creation has simply become unusual.
- 2.6 The significance of the relationship between humans, machines and jobs helps to set up a great debate. At one level it is quite simply an issue of whether we need to worry that robots will take our jobs, or whether we don't. This poses serious implications for humankind but to date the conversation has largely taken place on the fringes with many leaders ignoring the issue.
- 2.7 The conventional wisdom has been to minimise alarm and not to worry. Conversely, famous thinkers – including John Maynard Keynes, the most influential economist of the last 100 years, and Norbert Wiener, the acclaimed "father of cybernetics" – have suggested that the rise of Machine Intelligence leading to the collapse of human employment is a serious possibility (Cameron 2017:2). Between these two points of view lie a whole range of other perspectives including the reasonable expectation that is based on history and the disruption which has occurred as a result of earlier industrial shifts such as the Industrial Revolution and the more recent collapse of heavy manufacturing in the US and UK that led to the "rust belt," and a long and painful transition for the workers involved.
- 2.8 Of crucial significance is the fact that the effects of technological change are only felt when it intersects with the social context (shaped by responses from government and the legal system). Whilst we may not know the full picture, it

¹⁸² Smith, A., Anderson, J. 2014. AI, Robotics, and the Future of Jobs. Pew Research Center - Internet, Science & Tech. <http://www.pewinternet.org/2014/08/06/future-of-jobs/> [accessed 18th April 2016]

is likely that big disruption is coming to the labour market. How therefore are we preparing for these eventualities? (Cameron 2017:3). The full details may not be known but we might begin to consider the future by asking two key questions:

1. *On the conventional assumption that new jobs will emerge to take the place of those that go to machines, what kind of labor market turbulence can we expect during the transitions that will be involved?*
2. *Is the idea of big job losses that aren't compensated for with new jobs just ridiculous, or a serious possibility?*

2.9 Putting to one side for a moment discussion and involvement from technology experts and opinion leaders, there is no evidence of serious engagement with these questions by governments. One set of responses is that if the answer to Question 1 is "not much labor market turbulence," and to Question 2, "yes, the notion of big net job losses is just ridiculous," then there is no need to worry. Whilst these are both very comforting responses, they are unreasonable. And they are potentially dangerous, because they wrongly assess the risks we face going forward (Cameron 2017: 4-5).

2.10 Looking back over the collective wisdom of thinkers and opinion leaders, economists are near-unanimous in holding to the "conventional wisdom" that "while technology may displace workers in the short-term, it does not reduce employment over the long-term" (Cameron 2017: 6). However the Pew Research poll's result is dramatically different when experts are asked to look forward "signals the recognition that this wave of technological disruption could in fact be different. This cuts to the heart of the issue: *Is it different this time?* And if so, why might that be the case?"

2.11 In response, we have good reason to believe that, even assuming the conventional wisdom to be correct, we are likely to face substantial turbulence as careers and industries are disrupted right across the economy before the hoped-for "new jobs" emerge in sufficient numbers to maintain the full-employment norm (Cameron 2017:11).

2.12 Second, the possibility that this will not happen – that we shall instead see capital and technology incrementally substituting for human labor faster than new jobs can emerge – needs to be taken very seriously. It's a possible outcome that should be occupying our leaders; and our best thinkers should be addressing the question of how we might prepare. Here is a risk of the collapse of the "full-employment" norm to which all the developed economies have become used. It may be hard to estimate how great that risk is, but it is not trivial.

Care of the older person

2.13 The fastest growing population in developed nations is those aged 65 and older. It is estimated that there are currently 10 million over-65s in the United Kingdom - 1.5 million of those are over 85 - and the figures are

- expected to rise in the coming years.¹⁸³ Globally, over 60 year olds represent 11% of the world population and this figure is expected to double by 2050.¹⁸⁴
- 2.14 Within the next 20 years it is increasingly likely that robots will be used in the care of older adults throughout the developed world. This is a striking technological and social development with widespread but poorly understood implications for the society as a whole. It is critically important that the psychological, philosophical and spiritual implications are considered and debated before robotic care assistants become ubiquitous.¹⁸⁵
- 2.15 Of critical importance in the conversation is the anthropomorphic question and the extent to which the older person can connect with these machines. Do they simply see these devices as 'pets' or as 'human' in some way? In turning to AI devices to provide help and support are we helping to create a situation where those who are losing their mental and cognitive functioning will increasingly regard these devices anthropomorphically?
- 2.16 Dependency is a key issue in care of the elderly. Increasing in life span brings with it increasing levels of dependency. It is erroneous to suggest that living longer creates dependency, for humans are always dependent on each other. It is a feature of all our lives as we live interdependent, not independent lives. Increasing levels of dependency do prompt an increase in levels of intimacy both physically and psychologically. In the life of the older person boundaries of intimacy are being shifted, prompting fundamental questions: What can I control? What can I no longer keep private? What do I need to allow others to help me with?
- 2.17 In responding to these questions comes the possibility of abuse and objectification as well as any benefits of safety and care. As Bubeck notes, care is functional, involving "the meeting of needs of one person by another where face-to-face interaction between care and cared for is a crucial element of overall activity, and where the need is of such a nature that it cannot possibly be met by the person in need herself"¹⁸⁶.
- 2.18 In considering and assessing how robotics and AI assisted devices can benefit the older person, a functional relationship model needs to be adopted which requires looking at both sides of the coin: what does it do for the person being cared for and on the other: what does it mean for the one providing the care?
- 2.19 Care relationships also need to be considered in a wider social context in terms of what society will allow, what needs to be provided (social support, organization and administration) and what values should direct this.

¹⁸³ **BBC News Online. 2012. "Social care - how the system works" 10th July 2012,** <http://www.bbc.co.uk/news/health-18610954> [accessed 19th April 2016]

¹⁸⁴ UNFPA and HelpAge International, 2012 Ageing in the 21C: A Celebration and A Challenge <http://www.unfpa.org/webdav/site/global/shared/documents/publications/2012/UNFPA-Exec-Summary.pdf> [accessed 19th April 2016]

¹⁸⁵ Metzler, T.A., Barnes, S.J. 2013. Three dialogues concerning robots in elder care. *Nursing Philosophy*, Vol 15 (1) DOI: 10.1111/nup.12027 <http://onlinelibrary.wiley.com/doi/10.1111/nup.12027/abstract>

¹⁸⁶ Bubeck, D. 1995. *Care, Gender and Justice*. Oxford: Clarendon Press. p .129

2.20 Within this lexicon robots and AI devices hold the promise of being able to help facilitate relationships of care as instruments and tools within the overall sphere of care. Technology can assist with the *tasks* of care and we must ensure that it is not exploited for the *practice* of care.

3. The role of government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- 3.1 Governments and policymakers are faced with the challenge of how to respond to this 'disruption'. Conventional wisdom would indicate that we need only double down on innovation and the jobs issue will solve itself. Meanwhile, a growing cohort of smart and distinguished individuals is beginning to see things very differently, foreseeing a potentially major problem with fundamental implications for our social and economic assumptions about the future - and policy (Cameron 2017:12).
- 3.2 The best approach to help resolve this question appears to be in the form of risk. Policy makers need to synthesize these divergent possibilities into a single approach that is focused on risk. Let's look at all the plausible futures, decide how probable each one is, and do our best to prepare for all likely outcomes – and their respective risk profiles (Cameron 2017:58).
- 3.3 A simple approach to assessing risk starts by assigning possible outcomes to four categories – four boxes in a simple matrix: High Impact and High Probability; High Impact and Low Probability; Low Impact and High Probability; Low Impact and Low Probability. The first two are the ones that matter most. And the second is the trickiest.
- 3.4 In terms of work and employment, two scenarios appear ripe for the purpose of risk analysis. One refers to the likelihood of serious structural unemployment; the other to the likelihood of a failure to recover from such turbulence in the labor market and the long-term whittling away of the economy's capacity to sustain "full employment" in a future in which AI/robotics has been substantially deployed. Is there a High or Low Probability of serious structural unemployment? Either way, if it happens it will have a major impact on developed (and potentially developing) societies.

5 September 2017

Bioss International Ltd – Written evidence (AIC0033)

AI Governance

Where is the knowledge we have lost in information?
Where is the wisdom we have lost in knowledge?
– T.S. Eliot.

A Protocol for the Human/Machine Working Relationship

1. For many decades the work of Bioss has focussed on human judgement and decision-making and the conditions that leadership creates to support good judgement.
2. We are now concerned with the “working relationship” between human judgement and decision making on the one hand and machine judgement and decision-making on the other and with the implications of these working relationships for governance.
3. We see ‘good’ governance as being accountable to the providers of resource (investors, taxpayers) for the value created or lost by executive action. That executive action is realised through working relationships (until recently, solely between people and shaped by structure and culture). In current and likely future circumstances, creating and not losing value or leaving it latent will also depend on the working relationships between humans and intelligent machines.
4. We are suggesting a practical, non-technical, framework for managing a critical set of new working relationships with Artificial Intelligence, both now and in the decades to come.
5. We have chosen the word “protocol” because it implies guidelines that govern working relationships between two or more groups.
6. Clarity and boundaries are essential in human-to-human working relationships, especially in situations that are fast moving and unpredictable. They are equally essential for our working relationship with AI.
7. The concept of “vigilant trust” allows *people* in ‘shared-destiny’ relationships to define and experience trust in a way distinct from both the unquestioning faith of dependent relationships, and from the distrust of adversarial relationships.

8. We see the same tension between dependency and distrust emerging in the debate about AI. We at Bioss see AI neither as saviour of all things, nor as harbinger of doom. The reality is more likely to play out somewhat more messily, between “Machines of Loving Grace” as the Sixties counter culture would have had it and the more gloomy predictions of runaway AI deeming Homo Sapiens surplus to requirements.
9. We are however, whatever your perspective on where we are headed, *already* in a ‘shared destiny’ relationship with AI. The time for an “off switch” is over. There are too many actors, too many places where AI is being developed and is already deployed.
10. Rules that are too hard and fast will not keep pace with rapid change. Bioss therefore suggests asking one very practical and deceptively simple question - “What’s the work?”
11. Analyse the work that any given AI is doing, in any organisation in the public or private sector, at any given time. And be conscious and aware when certain key boundaries are crossed.

The Bioss AI Protocol & AI Governance

12. This core ‘checklist’ should be kept under permanent review **at the most senior governance level of any organisation** developing and deploying AI.
13. Each of the proposed ‘areas of inquiry’ has been carefully considered to encapsulate the key issues currently debated about AI and its impact on organisations and the wider society. They can sit independently and also in dynamic relationship to each other over time.
14. This review work can be done at the design phase, at the test phase and of course as part of regular review once the AI is ‘at work’. This review work should not be left to technical specialists and data scientists alone.
15. Multi-disciplinary review teams should also include key stakeholders (customers, citizens, ‘AI anthropologists,’ designated Board Members for example).
16. We suggest that this is a useful way of reframing questions about how AI will impact working relationships between humans, and what it will mean to manage or be managed by an algorithm, in part or in whole.

17. It is a different perspective on the ongoing debates about “Narrow, General and Super Intelligence.”

Bioss AI Protocol.

In relation to the human/machine working relationship, we would ask five fundamental questions about the work an AI (or a linked network of AI’s) is doing.

Each time consider how high the stakes are *for human beings* in the “work” the AI is doing in any given situation.

Is the work **Advisory** – leaving space for human judgement and decision-making? If so what assumptions lie behind the AI’s “advice? And whose assumptions?

Has the AI been granted any **Authority** – power over people or other machines?

Does the AI have **Agency** – the ability to act in a given environment without ‘recourse first to a human being?

How conscious are we – at every stage of AI deployment – about the skills *and responsibilities* we are at risk of **Abdicating**? What human skills will atrophy?

Are lines of **Accountability** clear? This is a critical issue and underpins each of “Advisory, Authority, Agency, and Abdication.”

For all the fallibility of human institutions, accountability lies with boards and governments. Can an AI *ever* be accountable if it cannot feel pain, responsibility, shame, and remorse or be meaningfully sanctioned?

18. If we are clear about the work being done and we manage the boundaries between the Five A’s with intelligence and compassion then AI will **augment** human judgement, creativity empathy and wisdom. If we don’t it won’t.

19. AI can and should be a fundamental part of the work we should be doing more broadly to create institutions that **maximise the constructive aspects of human nature and minimise the destructive**.

20. The AI Protocol is thus a simple but powerful analytic focussed on boundaries that we should cross consciously and deliberately as we embed AI more and more in our personal, professional and public lives.

21. It is a different way of framing some of the more abstract debates about AI and ethics (see below) by focussing on the every day practicalities for organisations in the public and private sector. And it is flexible, as working relationships will evolve in the years to come.
22. Each category permits a deeper set of analysis and inquiry, for example about the well discussed “bias” risk in the original data, about “explainable AI” or transparency, about who has accountability for the work that the AI is doing and at what level in the organisation.
23. Or for public policy in particular the “abdication” of whole swathes of work including “cognitive” work to AI requires careful planning in relation to training and the tax base on which modern welfare states depend.

More questions to ask:

24. When you “task” an algorithm with its work (be it with an advisory role or with real authority), are you fully conscious of the level of complexity at which it will be making decisions? Do you give it more “credence” than its current level of capability merits?
25. To what extent can you trust that algorithm to make decisions in line with strategy and purpose, with values, with an operating philosophy? What does it mean for an AI to be “trustworthy” if you have granted it significant agency?
26. As the organisation or government department’s strategy and purpose evolve to respond to changes in the wider context, what do we need to do to ensure that our algorithms evolve along with them? If AI takes significant roles within an organisation, or indeed in the wider society, which human jobs should be replaced and which new ones should emerge? What can be done so that societies can have that choice?

Ethics

27. This is clearly a very significant topic in its own right and it seems to us that there is a potential category error that we are making here.
28. We have not sorted out our own ethics as a human species. We are still not able to answer Plato’s question about the public good: whose good, and who decides? Nor are we clear that the answer would be a universal one, common across different cultures.

29. This is not to say that we should not be concerned with the ethics of AI. We should be, but we should also be wary of the magical thinking that there is an “ethical algorithm” out there waiting to be programmed, or waiting to emerge through deep learning – one that will know what the “ethical” thing to do in any given circumstance is.
30. Hence the more practical approach suggested by the Protocol. One that is based on vigilant trust, asks what is the work and determines how significant the potential ethical dilemmas may be for any given AI at work on behalf of an organisation in any given situation.
31. If the AI has been granted significant *agency* and *authority* for example then the review team working with the AI needs to be vigilant about those moments when the decision has significant moral implications indeed in the same way they should be with every key decision they make. This is a problem that is far older than AI but now includes it. It seems to us at Bioss that the American philosopher John Dewey and his work on ‘pragmatic ethics’ based on thoughtful, conscious, repeated and rigorous inquiry of specific ‘situations’ is of particular relevance to this field.

Intelligence

32. A second footnote about the debate some observers want to engage in over whether AI is ‘intelligent’ or not and the knotty issue of our inevitable impulse to anthropomorphise AI.
33. Douglas Adams, author of *The Hitchhiker’s Guide to the Galaxy* (featuring the supercomputer Deep Thought, which calculated the answer to the Ultimate Question – 42 – and Marvin the Paranoid Android, an AI burdened with existential angst), used to tell a mini-fable about a puddle.

"This is rather as if you imagine a puddle waking up one morning and thinking, 'This is an interesting world I find myself in – an interesting hole I find myself in – fits me rather neatly, doesn't it? In fact it fits me staggeringly well, must have been made to have me in it!' This is such a powerful idea that as the sun rises in the sky and the air heats up and as, gradually, the puddle gets smaller and smaller, frantically hanging on to the notion that everything's going to be all right, because this world was meant to have him in it, was built to have him in it; so the moment he disappears catches him rather by surprise. I think this may be something we need to be on the watch out for."

The Salmon of Doubt (2002)

34. This fable was Douglas' plea for a little more humility, when it comes to acknowledging that human cognition, perception and intelligence may not necessarily be the apogee. It was a characteristically witty and elegant challenge to the longstanding Aristotelian notion of the *scala naturae*, which runs from Gods and humans at the top of a ladder, downwards towards other mammals, birds, fish, insects and molluscs at the bottom.
35. Frans De Waal, in his book *Are We Smart Enough to Know How Smart Animals Are?* writes "cognition is the mental transformation of sensory input into knowledge about the environment, and the flexible application of this knowledge. While the term cognition refers to the process of doing this, intelligence refers more to the ability to do it successfully."
36. De Waal's book is about chimpanzees and capuchins, dolphins and elephants, corvids and octopodidae, but the working definition for cognition and intelligence holds good for AI too.
37. With Douglas Adams's puddle in mind, our working assumption is that these emerging silicon based (and one day quantum based) artificial intelligences may not be like ours, but that does not mean they are *not* forms of intelligence. Kevin Kelly in his book "The Inevitable" aptly calls them 'alien intelligences.'
38. This is a freeing idea. It is practical and does not overburden the work we need to do *now* in relation to AI with too much abstract principle, or issues of when and if AI will be sentient or even conscious. It also crucially allows for emergence and uncertainty. Embracing uncertainty is critical for managing our emerging relationship with AI.
39. The thing that all present-day AIs have in common is that they are *working for us*. (At least for the moment!) Philosophical conversations about whether artificial intelligence is intelligent, whether "they" will become sentient, and whether they will always have *Homo sapiens'* interests "at heart" are critically important. However, we need practical, accessible (not only technical, nor overly legalistic) frameworks for navigating this liminal space now.
40. One final observation about why we need to be vigilant about those systems that gain their "judgement and decision making capability" from data created by human patterns of behaviour and why we must have appropriate review mechanisms for the work we task AI to do, is well summed up by two quotations from the economist George Shackle, who writes about the essentially mutable quality of knowledge. The implications for AI, the work we give it to do and where and how it gains *its* knowledge are profound.

41. From Expectation, Enterprise and Profit:

"To say that there is always potential new knowledge to be gained is to say that possessed knowledge is always incomplete, unsure and potentially wrong." (2003)

42. From Business, Time and Thought: Selected Papers of G. L. S. Shackle:

"I shall here suggest only one of several semantic or logical objections. This one arises from the essential nature of the decision-making process in a context, like that of real life business, where knowledge is in the nature of things always incomplete because necessarily always being increased at one edge and eroded at the other and always being transformed and re-interpreted. At all times, some of our information is losing its relevance to decision because it refers to what is now an increasingly remote past. Its place is always being taken by fresh information relating to the more immediate past. And all the time fresh scientific discovery and fresh invention are rendering some knowledge obsolete or showing it to be false." (1988)

Conclusion

The Protocol is a practical way of consistently reviewing the changing relationship between human judgement and decision-making and machine judgement and decision making as the boundaries between the two evolve and shift in the weeks, months and years ahead.

29 August 2017

Dr Andrew Blick – Written evidence (AIC0064)

Evidence submission to House of Lord Select Committee on Artificial Intelligence

Submitted on personal basis

Dr. Andrew Blick, Senior Lecturer in Politics and Contemporary History, King's College London

Introduction

1. I am Director of History & Policy (H&P), a cross-UK academic network based jointly at King's College London and the University of Cambridge. It seeks to bring together historians and their work with policy makers, to the benefit of both. In the following evidence, submitted on a personal basis, I draw on a variety of projects with which I have been involved. They are: an H&P workstream entitled 'History and the Internet'; my own research on the history of constitutional reform proposals in the UK; and a pamphlet I am writing for The Constitution Society, jointly with Emily Barrett, on the Internet and the UK constitution. While the Internet is a distinct subject from artificial intelligence, there are important connections between the two.¹⁸⁷
2. The House of Lords Select Committee on Artificial Intelligence is a welcome initiative. An investigation in this area, building on the earlier work of the House of Commons Science and Technology Committee, is timely. As I discuss below, it is important that oversight of artificial intelligence should come from Parliament; and enhancements to the way in which representatives in Westminster hold government to account in this regard could be an important legacy for this Committee. I make my own proposals for such mechanisms, for the consideration of the Committee, in this paper.
3. This submission approaches artificial intelligence from an historical perspective, with particular reference to constitutional matters. As far as technological and associated issues are concerned, it takes a lay perspective, accepting conclusions presented in some of the secondary literature as a basis for the discussion of historical and constitutional considerations, but not interrogating the actual scientific and philosophical underpinnings.
4. I use the working definition of artificial intelligence provided by Transpolitica in written evidence to the Commons Science and Technology Committee, adopted by the Commons Committee for the

¹⁸⁷ In the words of the Government Office for Science report of 2016, *Artificial Intelligence: opportunities and implications for the future of decision making*, 'In the online world it is already a part of everyday life, sitting invisibly behind a wide range of search engines and online commerce sites.', p.4.

Robotics and Artificial Intelligence report it published in 2016 (Fifth Report of Session 2016-17, HC145).

-
- 5. It defines artificial intelligence as: 'a set of statistical tools and algorithms that combine to form, in part, intelligent software that specializes in a single area or task. This type of software is an evolving assemblage of technologies that enable computers to simulate elements of human behaviour such as learning, reasoning and classification.'
- 6. With this definition in mind, and the particular academic perspective suggested, I consider questions 8, 9, 10 and 11.

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

- 7. Here I focus on the democracy aspect of this question, and in particular the parliamentary accountability of the executive. Significant attention has already been devoted to the *legal* accountability of artificial intelligence. An associated issue, that should be of special importance to the Committee, seems to have been relatively neglected in public debate: that of *political* accountability. Artificial intelligence already plays a significant part in the way government takes decisions, which is projected to grow further. Such a development has major democratic implications. The government appears to be aware of this issue, but it is nonetheless important to rehearse it, especially from the perspective of parliamentary accountability, and ask how it might be addressed.
- 8. History suggests that changes in the nature of government create pressure for alterations in the means by which it is held to account. For instance, during the twentieth century, the idea of central government taking on a more socially interventionist role to enable it to tackle complex problems prompted calls for an accompanying adjustment in the way Parliament operated. A prolonged debate ensued in which various constitutional models were proposed, some of which were intended to facilitate more thorough oversight of the executive in changed circumstances. One important outcome was the House of Commons departmental select committee system as introduced in 1979. Though it had precursors, this reform – coupled with developments in the Lords committee system – has arguably brought about a qualitative shift in the way in which Parliament achieves executive accountability, which can be traced in some ways to discussions commencing in the interwar period about the expansion of governmental activity and its consequences.

9. Artificial intelligence could – and should – prompt a similar debate leading to new parliamentary practices.
10. From one perspective, if artificial intelligence can lead to the more effective delivery of services required by the public, it is desirable from a democratic perspective. But it could raise difficult questions about a crucial constitutional doctrine that can already at times seem nebulous: that of ministerial responsibility to Parliament. This principle is crucial to the working of representative democracy, since it is the means by which Parliament, including its elected component, the Commons, can oversee the executive on behalf of voters. The doctrine holds that secretaries of state and their equivalents are individually answerable to the legislature for their policies and decisions and the activities of their departments.
11. But if artificial intelligence comes to play an increasingly important role in Whitehall, political accountability problems may develop. Potentially, artificial intelligence systems that are learning and changing on their own account, becoming more autonomous, could render the idea of ministerial control less meaningful. Yet ministerial control is central to the idea of individual ministerial accountability to Parliament, that in turn lies at the centre of our democratic system. Moreover, artificial intelligence systems are known sometimes to reach decisions for reasons that are unclear to the outside observer. Such a tendency manifested within government could entail another political accountability problem. The ability to interrogate the rationale underlying a course of action is crucial to meaningful accountability. If a minister is unable to explain to Parliament and the wider public why a particular decision was made, democracy may to this extent be compromised.
12. In circumstances of defective accountability, the public may begin to question whether artificial intelligence is functioning properly and fairly.
13. The government accepts the existence of some of these potential problems and has introduced principles and procedures intended to avoid their emergence. For instance, in the 2016 paper *Artificial Intelligence: opportunities and implications for the future of decision making*, the Government Office for Science recognised that, given the need for accountability, special rules must apply to government. One reassurance it offered was that it was (p.10) '[l]ikely that many types of government decisions will be deemed unsuitable to be handed over entirely to artificial intelligence systems. There will always be a "human in the loop".' But the paper also noted that '[t]his person's role...is not straightforward. If they never question the advice of the machine, the decision has *de facto* become automatic and they offer no oversight. If they question the advice they receive, however, they may be thought reckless, more so if events show their decision to be poor.'

14. The 2016 Government Office for Science paper also referred to the regulatory structure provided by the *EU General Data Protection Regulation 2016* and the *Data Protection Act 1998*.
15. A further document, the 2016 *Data Science Ethical Framework*, issued by the Cabinet Office, sets out six 'key principles' applying to the use of data for decision-making purposes within government. Alongside other matters, they address democratic concerns. For instance, principle two requires 'minimum intrusion'; principle four calls for an awareness of 'public perceptions'; and principle five requires users to '[b]e as open and accountable as possible'.
16. The principles that the government advances are clearly valuable and well-motivated. However, the executive should not be wholly responsible for determining the ethical regulations that apply to it, or for enforcing those rules once devised. While some laws operate in the area, potentially providing the judiciary with an adjudicatory role, there is also a need for continual, autonomous political oversight, that must come from Parliament.
17. How might full parliamentary oversight of the use of artificial intelligence be achieved? A minimalist approach would be to make this task a specific responsibility of Commons select committees, perhaps with extra staff and technical support made available to support this function via the Commons Scrutiny Unit.
18. But, more ambitiously, perhaps reflecting the possible scale of the challenge, there could be value in adapting the approach taken by the Commons Committee of Public Accounts, supported by the National Audit Office. A parliamentary committee for Artificial Intelligence Oversight (AIO), or perhaps an agency reporting to Parliament, suitably resourced, could be established. Assuming it is a parliamentary committee, it might be a Joint Committee of both Houses, taking evidence and issuing reports to inform Parliament. The work of the AIO ought to focus on the implementation of policy, rather than its merits. It would be entrusted with monitoring across government whether artificial intelligence was operating in accordance with the policy objectives it was directed towards, and was doing so effectively, and in accordance with prescribed norms.
19. Admittedly, the performance of such tasks would presumably involve a degree of technical innovation, but such is the nature of the field. Moreover, for the AIO committee properly to carry out its functions, it would be necessary to establish terms of engagement with the executive. It might be worth establishing within Whitehall new a role equivalent to the accounting officer. Its holder would have special powers in relation to the signing off of the artificial intelligence policies of departments and associated agencies, and answering personally to Parliament, rather than on behalf of the secretary of state.
20. The AIO committee could take on ownership of a revised *Data Science Ethical Framework*, which would include within it stipulations regarding

the way in which this new system of oversight functioned, as well as regulations applying to the use of data for decision-making.

21. This new system might be given statutory force.
22. It might further be desirable to develop specialised ombudsman functions, enabling members of the public to seek redress in relation to the operation of artificial intelligence where they feel it has been inappropriate in its impact upon them. This function could be attached to the AIO committee, or located elsewhere.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

23. In the public sector, the default assumption should be for transparency. Public authorities, in the broadest sense – that is any body, whether publicly or privately owned, exercising a public function – might be subject to rules akin to those in force under the *Freedom of Information Act 2000*. The expectation should be for transparency, unless the activity concerned falls within one of a number of defined areas for which confidentiality is necessary (such as security and intelligence, defence, law enforcement, and market sensitive activities). Beyond the public sector, it might be appropriate positively to define those areas in which transparency is required. They could include activities in which it was possible that the use of artificial intelligence could lead to discriminatory outcomes and therefore the compromising of individual rights.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

24. The government should bring forward draft proposals for a comprehensive regulatory framework for artificial intelligence in both public and private sectors; alongside a long-term policy strategy statement. This process would lead ultimately to an Artificial Intelligence Act. It would lay down requirements pertaining to the protection of individual rights such as privacy, non-discrimination and not suffering on a basis of anticipated future actions. It would introduce stipulations around the transparency of artificial intelligence systems. The Act would also create a statutory basis for the parliamentary AIO committee, underpinning its terms of reference, resources and independence from government. It would create a statutory footing for the revised *Data Science Ethical Framework*. Finally, the Act would provide for cooperation with other countries and international organisations in the regulation of artificial intelligence. While the current government preoccupation with exiting the European Union

may be a distraction from such a major undertaking, it is no less urgent. Indeed, since proper regulation of artificial intelligence would seem most appropriately handled at supranational level, it becomes all the more important to ensure that the necessary collaboration can be achieved in a post-EU future.

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

25. One apparently worthwhile model exists in the form of the EU General Data Protection Regulation, referred to above, which will come into force in 2018. The Regulation will establish the principle that individuals can request an account of why a particular decision was reached in relation to them by an automatic system. The UK will not benefit from the applicability of this regulation once it is outside the EU, and may experience difficulties if it seeks fully to replicate it within its own domestic legal system. This issue is of course part of a larger challenge relating to the status of former EU law post-Brexit.

4 September 2017

Dr Paula Boddington – Written evidence (AIC0067)

1. Artificial intelligence can be defined in different ways which will have an impact upon how the ethical issues are perceived and addressed. One of the key ethical aspects of AI is its use to enhance, replace or supplement human agency, either individually or in social systems. I consider that keeping attention on these aspects of AI is essential for addressing its major ethical challenges. This will impact upon the methodologies we use for addressing the issues.

2. Is the current level of excitement which surrounds artificial intelligence warranted? Whatever the reality, the hype around AI, both concerning its technical capabilities, and the moral implications, should be a topic of direct concern, since it can severely distort our thinking about the search for ethical ways forward. It is vital in thinking about the ethical implications of AI, that on a case by case basis, we consider what elements of AI present new problems, and what elements of AI are continuations of familiar issues.

3. How can the general public best be prepared for more widespread use of artificial intelligence? This question is phrased in ways which make the general public sound like passive consumers of a coming revolution in which they have no say. THIS is precisely one of the major problems. It is vital that the general public have as much education as possible about how AI is operating in their lives, how it affects them, and how to ameliorate or avoid aspects of this that they find objectionable. It is vital that the general public does not find itself nudged into a situation where large effective monopolies, such as search engines etc, are using AI in ways which have profound reach over work, education, and social life, over which ordinary citizens are powerless, or which an individual can opt out of only by forgoing involvement in key aspects of life. See, for instance, how important Facebook has become for networking in certain careers.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated? There is much talk about how AI will change the workforce, and how certain jobs will be eliminated. Much of the excitement around AI seems to be around pushing humans beyond their top limits of our species' intellectual capacity, with the consequent dangers that a large section of the population may be 'surplus to requirements' (in a nutshell). It is already the case that those with lower intellectual skills have increasing difficulty in entering the workforce. I would strongly suggest that attention needs to be given to developing AI that might boost the capacity of those with lower intellectual skills if this is at all possible, who are after all, on our universalist ethic of human dignity, every bit as valuable as someone with an IQ in the genius range. The focus on an explicit or implicit 'transhumanism', i.e., a focus on busting through the ceiling of current human capacity, appears to have overlooked this possibility in many respects, although there are promising efforts e.g. to enable communication for people with various brain injuries, and the use

of robotics in teaching children with autism. I would strongly hope that the government supports as many initiatives focused on these and other less advantaged population groups. There are many changes in society as a result of the use of AI. One of the biggest questions is how to determine if these changes count as 'gains' or 'losses'. This very question should be addressed head on.

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how? In other areas of rapidly developing technologies or professional practice, such as in medicine and biotechnology, there has been great societal benefit from the presence of groups such as patient support groups and other independent means for the wider public to input their views, and often to act as watchdogs and to keep up pressures on professionals. Such is to be encouraged in the case of AI, keeping in mind that the organisation and membership of such groups into a number of sometimes disparate and sometimes competing 'publics' should also be noted.

6. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy? The IEEE Standards Association has, as you are no doubt aware, various working groups examining ways of producing standards to encourage beneficial AI. The committee P7003 is currently engaged in working towards developing standards to combat bias in algorithms, one important aspect of the many ethical issues around data-based monopolies. See <http://sites.ieee.org/sagroups-7003/>

7. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? This is a very broad question which cannot be answered simply. As stated earlier, a key aspect of AI for understanding its ethical implications is that in general, it aims to enhance, supplement or replace human agency. There will be knock on effects of this in the wider social setting including society as a whole, and including those who do not directly participate in the AI. As with the development of any technology, the population can end up accepting changes over the course of time which they would not have accepted, had these been implemented over a short period of time. This may or may not indicate a problem. What it does indicate, is the need for historians working in these areas to track precisely what changes are taking place. This is especially the case with AI, given that it is embedded, often invisibly, in so many areas of life. Of particular concern is how dependence on various forms of AI may impact upon education and on how our brains and bodies even develop and function, in turn affecting how we assess the ethical impact of AI. I strongly recommend that the government funds and encourages research into these aspects of AI.

8. I consider that the general form of codes of professional ethics assumes that the professional has an adequate level of control over their products and services. The problem of control and of transparency in AI makes such an

assumption problematic, hence, there should be no complacency about the magnitude of the problem with which AI presents us. Moreover, the standard normative ethical theories which philosophers and others currently draw upon in discussing ethics, tend either to bypass issues of human agency, or to make assumptions about human agency which are rendered problematic by the development of AI. Again then, we should not underestimate the scale of the ethical challenges awaiting us. I discuss these issues, including some common pitfalls in considering the ethics of AI, in my forthcoming book (see below).

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible? A key aspect of ethics and of ethical conduct is the responsibility to provide explanations and justification for conduct to relevant others: this is a form of transparency. Given the lack of transparency of much of AI, this responsibility can only be emphasised wherever possible. If there are deleterious outcomes for certain individuals or groups as a result of the operation of black box AI, this presents a serious ethical problem; one that cannot be overemphasised. The basis for professional codes of conduct includes the assumption that professionals have obligations towards clients and the public, based upon their ability to control the impacts of the products or services they produce. In AI, this foundational principle is in jeopardy, because of the black box and control problems. Every effort should be made to address this issue – probably on a case by case basis. For instance, in some issues, legal regimes of strict liability may be a way forward. There are many other instances where procedural transparency is a key aspect of maintaining ethical standards. For instance, in the courts, justice must not just be done, it must be seen to be done. 'Black boxing' would hence be unacceptable. The dangers are that as a population, we might be lured into accepting it in more and more instances.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how? The government could have a useful role in funding research in AI in areas that might affect or benefit neglected or disadvantaged population groups, where there are gaps in current research. AI exists in so many forms and with so many applications that a blanket 'regulation' for AI is perhaps unrealistic. Key elements that any regulation needs to consider are the problems with control and transparency. Further important issues concerns the general impacts on society, including on those who have not adopted any particular form of AI, and careful consideration should be given as to whether regulation in such areas is possible or desirable, or whether some other ways of monitoring and addressing concerns should be considered or implemented. There are frequent calls to address the possible impact of greatly increased unemployment that may arise from the use of AI, by a universal basic income (UBI). I would like to note that whatever the pluses or minuses of such a policy, one great concern I have is that this would increase the power that governments have over the lives of individuals, and could decrease individual autonomy – ironic given the use of AI that itself operates autonomously. I would urge the

government to consider this carefully and to respond with caution to those who hail a future where vast swathes of the population are reduced to a life of 'leisure' on UBI. Such is the stuff of totalitarian nightmares.

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence? The IEEE's Global Initiative has many admirable aspects, including the way in which it is focused upon attempting to incorporate ethical considerations into concrete standards for engineers, and the way in which it is attempting to have wide participation in its discussions. It is also important to learn from other fields where technology is impacting upon how humans view themselves, such as medicine, genomics, and biotechnology in general.

12. I discuss many of these points at greater length in my book, *Towards a Code of Ethics for Artificial Intelligence*, Springer, 2017 (to be published this October). <https://www.amazon.com/Towards-Ethics-Artificial-Intelligence-Research/dp/3319606476>

13. I have been working since 2015 at the Department of Computer Science, with Professor Mike Wooldridge and Professor Peter Millican, on a project funded by the Future of Life Institute with a grant from Elon Musk and the Open Philanthropy project. These comments represent my personal views and I am not here commenting in any official or representative capacity.

Dr Paula Boddington

4 September 2017

Michael Borgeaud – Written evidence (AIC0233)

There is a connection between certain interests I had when I was a Principal Lecturer in Sociology and certain issues raised briefly in the recent Oral Evidence Sessions. The key issue I shall comment on is the potency of Artificial General Intelligence, which in a Session was called “the big dream”. Before my specific comments, there are some background points that could clarify my underlying perspective on this topic that follows. It is not forwarded as a superior perspective. Instead I apply this type of paradigm or foundation (not a philosophy) because it seems to be a novelty here.

In the background, or as a typical side issue that exists in many technological and even scientific debates, some key words are understood differently. This is not a criticism. The point is that many of the disagreements as to benefits or problems with Artificial Intelligence might stem from different understandings of key words. There have been all sorts of debates and disputes about AI, over years and across nations, plus many dissimilar descriptions and definitions as a result of many technical developments. Likewise there are all sorts of names for types of AI, such as: Full, Narrow, Strong, Weak, Specialised, Assistive, Human-level, Superintelligence, etc. Even though in the new Sessions the various comments are very well informed and fluent, they are not easily amalgamated for some listeners who prefer a standard terminology referring to AI, as well as robotics.

Besides those different interpretations there are different predictions as to when and what types of AI will come into practice. You might find guesses that differ from 5 to 500 years, and beliefs that it is unpredictable. There has also been a huge number of publications. It is not surprising when experts say they will never manage to read them all. Another complication is that many components of AI have not been clearly explained. So there are uncertainties expressed, as well as the incompatible points of view - some very sceptical and others very optimistic. These ambiguities are not unusual considering how there are very different goals in the actual construction of AI machines. The diverse ways people talk, write and read about the technology is not quite the same as the actual work.

Likewise what I am writing can be regarded as another interpretation of interpretations.

The developments of the various types of AI (and as noted, having different names) are said to range from machines carrying out one straightforward specific job detached from human thought (a type probably named “Weak”) and at the more ambitious far end is the “human-equivalent” Artificial General Intelligence (AGI). Optimistic predictors, who reject the fearful pessimistic notions, insist that this system will be able to: reason as well as and with humans; explain how humans think; evolve styles equivalent to self-

consciousness; make decisions in various settings; and even surpass all our capabilities. These ideas rely on certain academic accounts of how people do a job, indeed do anything. Such accounts should be treated as theories about human behaviour rather than facts.

What I'm inclined to treat as a valid challenge to these predictions is not simply pessimism, moral disapproval or fear of the technical expenditure. Instead there are quite different foundations from which to conceive humanity and how we make sense of reality. I was encouraged to focus on this alternative view after hearing what Lord Giddens said during the 10th October Oral Evidence Session:

"To be able to master meaning you have to be in the real world; you have to be an agent; you have to have a saturated knowledge of human society. I just don't see how AI could ever get to that".

Such ideas could also draw attention to the way the orthodox and ambitious conceptions of AGI depend on a narrow conception of "intelligence". This also relies on narrow versions of Psychology.

AGI technicians seem to deal only with human brains, minds, cognition, reasoning, maybe emotions. A recommendation, related to Giddens' view, is that there needs to be new research organised with Sociology. It should be noted that there is not just one School of Sociology. And there are different approaches to AI within that discipline. The same complexity of varieties applies to Psychology.

So my approach to AGI in particular is shaped by particular sociological outlooks. It is intending to go into greater details than were necessary in those inspiring comments in the Session. Also, this approach is outside the more common frameworks of modern arguments for or against the chances of computers relating to people. I am not backing the dread that AI will be dangerous. I am more in favour of the view that AGI can *never be equivalent*, let alone superior, to men and women; a view that rejects the idea that eventually computers will manage all human things nicely. Another style of analysis that I reject assumes that, not some, but most people display the basic internal nature of being irrational, unwise, illogical, delusional and living in fantasy. Hence AGI is treated as progress. These ideas rests only on versions of "human science" that are thought to reveal our entire true nature and thus how to imitate or improve it.

These AI experts seem to rely on 'facts' about inner human 'things' such as consciousness, the brain, perceptions and intuition. Such specialised, yet changing, fields provide valid careers for sincere, wise people. Even so, they can only deal with fragments. Some critics say these specialisms often reduce what the critics call the subjective, enigmatic, ineffable infinity of humanity to isolated bits or 'factors' that get named, measured numerically and then put on the agenda of the AI technicians.

For ages it has been impossible to have practiced every type of science. So in this modern position AI technology must use only one of many scientific methods. Even with a specific scientific approach, its findings about human

behaviour are not unanimous and often impermanent. The fundamental mechanisms and roots of actual behaviours seem inexplicable. Some specialists admit they may never unravel such mysteries, let alone devise a computer to handle all forms of our behaviour.

Be that as it may, some specialist versions of 'human science' involved with the expansion of AGI to connect intimately with people, assume they deal with the reality of human behaviour. It seems that the plans for creating AGI are based on idealised interpretations of how *individuals* perform things.

A model of a rational actor is relied on when designing the special computer.

Discord arises because the many broad concepts of 'intelligence' have been explored in diverse frameworks, from nature to nurture, genetics to culture, individualism to national histories, etc.

Exactly what parts of these unconnected abstractions can or should be materialised into high-tech machines is unknown. The lack of agreed analyses results in varieties of artificial intelligence that are simplified, fractional imitations. Others say these are necessarily quite different to intelligence.

Over and above these disputes is the claim that AI mechanism cannot initiate their own agendas; they have to be programmed to begin with. There are varied theories as to how far programs are still needed to enable a machine to generate or improvise its response to external data, to achieve goals 'sensibly' in ways comparable to human minds.

But the Lord Giddens sentence implied that there are alternative perspectives of people, one of which deals with collectivist cultures not just individuals. It sees how members of a *community*, facing a goal, move through *situated* actions, often improvised, taking into account the current know-how of that location. Members' methods are not just rational reasoning, but part of the lives and everyday surroundings shared with others. Thus the formal, conventional rules for activities (presumably data to be built into AGI) are not the whole story. People make use of those 'official' criteria and formal standards only as part of their own pragmatic, essential methods, rarely talked about. Of course we often enact rules, repetitive routines and mundane traditions; and often act solo. Styles that might suit AGI. However, sceptics have emphasised the diversity of our natures, often dealing unpredictably and creatively with all sorts of new situations: doing odd extra things off the record, with tricks of the trade, rules of thumb, and folk wisdom. Alternatively sometimes people do not live up to expectations.

The acts are not simply kinds of intelligence. No matter how 'intelligent' and swift an AGI gadget might be, it cannot be prepared to grasp such common, situational and instant, spontaneous variations. Nor can it 'see the big picture', know how 'the world works' as we do, or 'think out of the box' and innovate in 'practical-moral' ways equivalent to humans. AGI based on digital computers must follow the rules of its program. Those rules are deterministic and therefore

do not promote the equivalent of genuine, spontaneous free will. Even if an AGI computer is hard-coded with billions of records of peoples' acts so far, for it to apply, that is not reliable to fit in with every next situation. Thus there will not emerge a device that can beat an average human in everything he or she does.

Now here is a major challenge to AGI – a depiction of the *mind*. It is not to be seen as an internal organ, or an immaterial 'spiritual' element. Instead it is a complex, every day, *social* phenomenon; a way of doing things. Of course there are times when a person on their own gets occupied with a silent, private 'inner self' usually 'felt' to be inside the head (or body). Even so such everyday events are still part of our culture's settings. Those scholars I prefer contend that all that our heads contain is complex, physical brain materials - organic, biological and chemical. Their premise that minds and social life are conjoined means that all things we do are the products of the indivisible personal AND situational influences.

So the way brains work is not the essence of *language in use*. That should be treated as witnessable activity rather than cognitively generated. Anatomical versions of brain functions do not provide AGI with reliable versions of our regular conversational techniques. Most designers of AGI use orthodox human science versions of the 'machinery' of language so as to model natural *talking*. Sometimes collections of people's talk get converted into digital data to be uploaded. Of course some of our chats are mundane repeats, commonplace and systematic. People cannot say anything anywhere and mean anything. Talking is orderly not haphazard. Languages are patterns of sensory meaning, and conversations are more than patterns of words.

Apart from some conversation patterns that might be theorised, there is an analysis that treats each record of people's talks as unique, contextual and functioning instantly. The way words are applied, what they are intended to mean, how they are heard and understood, depends on each specific situation. People show a vast array of linguistic styles and pronunciation. We share a set of words which are used in action to express all sorts of requests, complaints, commands, promises, warnings, questions, apologies, excuses, jokes, lies, memories, sadness, sarcasm, anger, etc. Speaking includes interwoven 'variables' such as loudness, pace, rhythm, tone, pauses, hesitations, interruptions, etc. Also, to convey what we mean does not always need a completed speech. Vitally intrinsic with talks are non-verbal communications, such as facial expressions, glances, eye contact, body movements, nods and gestures. Inseparable from how talks work is listening, replying, not just understanding or agreeing. Gaze direction such as not looking at the speaker effects the other's behaviour and replies.

Equally inherent is what speakers are *doing* when they talk. The conversation and attached acts can be spontaneous or casual, planned or unplanned. These circumstances can be not exactly like anything before. There can be problematic

features as there are in everyday life. However, quite often people say novel things in novel ways that work. Another feature of real talk that does not fit well into AGI, is how and when it is understood. That does not happen at a standard level after each sentence, but develops variably, unpredictably, yet usually towards the ending of a conversation. Also whilst chatting, a split second pause (not easily copied by a computer) can be consequential.

Thus: the meanings of talks and their consequences are in movement, progressing in a mobile context that is not that predictable. Not only words to each other on specific topics, but all interactions in real life cannot be accurately reduced to *mechanistic* programs. Humans do not operate on computational principles, so there is no overlap with AI. The super-computers cannot duplicate our socially formed natural ability to know instantly what anything is, and to handle it. Understanding is a predicate usable for persons socially, and not simply complete in their 'minds' or their brains.

How can AGI programming convert these phenomena into an electrical, empirical equivalent? Computers are not actually "hearing" anything like we do. We are not just hearing 'sounds' in conversations, it is what speakers are doing, meaning, hinting, or wanting. These rules are never absolute or complete. For example, instead of merely responding to a suggestion, people may turn their response into a mock tease. How could machines be trained to argue, tease, or exaggerate?

More realistic is to state the obvious – that ordinary commonsense is the foundation of everyone's everyday activities. It is workable, fallible, variable and inseparable from the fluid range of mobile, combined surroundings, for example – what counts as the relevant environment, other people's attitudes, the time being spent, the relevant historical and cultural factors as well as all sorts of actual situational things that could even include the time of day and the weather.

CONCLUSIONS

People's experiences can only be understood in context.

Humans might not be able to process data as fast as computers, but they can think abstractly, plan and solve all sorts of problems at a general level without going into the details.

They can innovate, come up with ideas that have no precedence.

Technicians and other people often treat these machines as if they were humans, which is pleasant and understandable. However the main issue that researchers admit, is that the machines may imitate lifelike behaviour but they are not alive. That is the fundamental but avoided basic fact.

Michael Borgeaud – Written evidence (AIC0233)

AI devices will never be able to talk to people as we talk to each other.

Human creativity, subjective judgement, and everyday craftsmanship will remain beyond any skill a machine can offer.

5 December 2017

Braintree – Written evidence (AIC0074)

Braintree

House of Lords Artificial Intelligence Committee Submission

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1 Artificial Intelligence (AI) is currently undergoing a renaissance. Improved technologies and rapidly increased connectivity have put the potential of AI at the forefront of innovation once again. With advances in our understanding and capability, this much talked about new technological frontier, previously unachievable, is now within our grasp. The way that AI develops, in the UK, over the next 5, 10, 20 years is going to be contingent on a number of technical and societal factors that could all either accelerate or hinder its development. A key accelerator/barrier that Braintree is specifically trying to address through its work is the processing of connected data to useful effect.

1.2 Individuals, companies, governments and services are creating more and more data at an exponential rate. The potential this data presents is hugely significant. However, our ability to process this data, and relate it to other sets of data, is also becoming exponentially more complex. Our smartphones, tablets, computers and laptops are growing ever more powerful, and thereby consuming ever more amounts of data. They are also producing more and more complex data. Our ability to connect more and more to each other and the appliances around us is increasing – but this rate of connection is growing at a rate that neither our infrastructure or current technologies can really keep up with.

1.3 There are already plenty of examples of where technological formats, operating systems, and so on, will not 'talk' to each other as there are too many formats to handle. We as individuals are, very soon, going to need 'universal' technologies that can aggregate all these different data streams and relate them to each other. That is where AI has the real potential (and urgent need) to make sense of all the data that society is producing. Data in a vacuum is largely useless. The useful information in data is often like looking for a needle in a haystack, if you don't have the right technology to look for it. Indeed, the pace of data we are creating is causing a situation where we are looking for different needles in a variety of different sized haystacks.

1.4 But this is where AI can do things that humans (or even most conventional computing) cannot do. AI solutions like those designed and championed by

Braintree, have the ability to almost instantly compare, contrast, and relate disparate data sets, and then apply the learning to useful effect. In fact it can sometimes find things you were not even looking for.

1.5 We were recently working with an international retailer to examine their metadata to explore opportunities for making their promotions schemes more effective. What we ended up finding was a multi-million pound fraud through misallocated discounts and vouchers (the company in question were entirely unaware that this was taking place. Our AI's ability to compare and relate the data of millions of promotions, millions of customers, millions of transactions was able to discern patterns that exposed the discrepancies.

1.6 The relevance of this example is two fold. Firstly, this is just one example of the powerful insight AI can provide when applied to large datasets (of essentially human interactions). Applied at scale to public datasets the potential could be enormous. Imagine the patterns that intuitive AI could draw in our nation's health data, our travel data, our pensions, benefits, even our overall public spending?

1.7 Secondly, and importantly, the above example is useful as it bears out the core component required to make AI effective – and that is large data sets. AIs, of the type that Braintree work with, use a technology known as 'dynamic graph databases'. This process allows for large scale data sets to be compiled, merged, related and analysed at great speed. The AI then produces reports on the relationships in the data. You of course still need professionals to interpret the findings to establish the usefulness of what has been found and propose solutions based on the evidence. Though not initially, there is even the possibility for these professionals to interact with the AI using spoken language. The key is that the AI is able to digest the data rapidly, at scale, and discover relationships that human researchers, or other computing techniques, are simply unable to do.

1.8 The technical ability to compile and mine data is absolutely there. However, for these technologies to succeed there needs to be societal buy-in. It is of course largely individual data that makes up the useful large scale data that AI can interrogate. Therefore, the public need to be comfortable with its use; and here lies one of the key challenges to the successful implementation of AI. The public and state needs to be comfortable with responsibly opening up data for interrogation and analysis. This requires overcoming a significant perception barrier of AI and data use. Indeed, much of the public and governmental suspicion of AI and its role stems from misconceptions of its capabilities. People are understandably nervous of the concept of an 'unaccountable' AI having access to their personal details, and using the data in ways that the analysts do not fully comprehend. Many are suspicious that their details are processed in a 'black box' that cannot be fully interrogated.

1.9 However, this is a misconception of how many AIs work. AI has become a catch all term for a wide range of technological capabilities. Not all, indeed very few, operate through a 'black box' 'deep learning' model. The AI that Braintree itself pioneers works through 'neural pathways' – in practical terms this is different strands of computing tacking individual seams of data, all of which can be separately interrogated, and therefore the thinking behind the AIs conclusions can be reverse engineered.

1.10 To give an example of this, one experiment we recently assisted was for an airline in developing an AI-led auto-pilot system. This system was able to land a simulated plane in exceptionally adverse simulated weather conditions. If our AI system was a 'black box' method it would have been very difficult to explain how it had achieved a safe landing. Not being able to explain the exact decision making that had gone into the descent would obviously be a significant barrier to any airline wanting to use the technology, especially when you're entrusting it with the lives of hundreds of passengers. Instead, our programmes are able to precisely explain the minutiae of each flight decision, and therefore instil significantly more confidence in the system.

1.11 Where AI has the opportunity to benefit a number of sectors is in the area of 'modelling,' in a business sense. AIs will increasingly give businesses and policy makers the ability to model their ideas and plans before implementation. This modelling can be applied in any processes: social life, economics, medical research, city development etc. Effective use of AI-led computer modelling will increasingly set apart successful businesses and government in future.

1.12 Fundamentally it is up to the UK how it wants AI to develop, and how it wants to lead in this area. AI technologies will develop apace, with or without government involvement. Other country's governments will develop their own AI priorities and technologies and channel it as they see fit. Therefore, if this country wants to be at the forefront of responsible AI development, then it must lead from the front – AI won't wait for the UK to catch up. Government and industry must now rapidly work together to develop 'terms of engagement' to set the parameters of responsible and ethical AI innovation in this country.

2.2 Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

2.1 When discussing the current development and deployment of Artificial Intelligence, it is important to differentiate between 'strong AI' and 'weak AI'. Weak AI, at this moment in time, is becoming more and more commonplace. A consumer in the digital world may encounter weak AI frequently, though they may not think of it as 'artificial intelligence'. One example is the "related products" service that you may find when shopping online. This weak AI system

is able to, in a rudimentary fashion, find products that are in some way related to the product that we are viewing or purchasing and present them to the customer. This is weak in the sense it is not genuine intelligence, it cannot apply solutions to a range of problems. It can only operate within a pre-defined range of rules set upon it towards one specific task, rather than create *new* solutions.

2.2 When it comes to weak AI, there are some small scale societal benefits, and some cosmetic. Weak AI may enable us to access our data more efficiently, it may help recognise our preferences when it comes to the temperature of our homes or our chosen seat position in our cars. Our appliances and gadgets may get a bit smarter, our manufacturing may get a bit more efficient, and processes that used to take large amounts of manpower now may take less. However, the benefit of such AI usually rests with those who own or can access the technology – it would not be fitting to describe the technology as particularly transformative. However, on the other hand, strong AI - the AI that companies like Braintree are developing - has not been fully realised anywhere in our society today. Strong AI, the development of machine processes which can make its own decisions skilfully and flexibly, has the potential to completely transform our society and, if done right, can make Britain a truly global leader.

2.3 However, AI is often discussed in the context of benefitting corporations or governments, with little regard for the benefit to every day members of the public. It is talked about as a high level, at times abstract, concept used for making efficiencies which, to the individual, can be interpreted simply as job losses. Indeed, AI as a subject is rarely broached in the media without references to impending doom in the jobs market¹⁸⁸, with the work of whole industries being made redundant thanks to the development of a few machines. Internationally, we have seen how some nations are seeking to combat this perceived threat¹⁸⁹.

2.4 It is important to make clear here that AI is not robots, and AI is not automation. Though they are related, they are separate entities. All major manufacturing industries use robots and automation right now, without the presence of AI. We can, right now, build millions of robots without any AI.

188 'Robots will take a third of British jobs by 2030, report says'

<http://www.telegraph.co.uk/technology/2017/03/24/robots-will-take-third-british-jobs-2030-report-says/>

'40% of jobs' taken by robots by 2030 but AI companies say they're here to help'

<http://metro.co.uk/2017/05/10/40-of-jobs-taken-by-robots-by-2030-but-ai-companies-say-theyre-here-to-help-6628469/>

'Millions of UK workers at risk of being replaced by robots, study says'

<https://www.theguardian.com/technology/2017/mar/24/millions-uk-workers-risk-replaced-robots-study-warns>

189 'South Korea introduces world's first robot tax'

<http://www.telegraph.co.uk/technology/2017/08/09/south-korea-introduces-worlds-first-robot-tax/>

2.5 When it comes to AI specifically, there is likely to be changes in the job market upon its widespread implementation. However, what scare stories of mass unemployment fail to take into consideration is the jobs that will be *created* by this technology, jobs that we may not at this stage be able to comprehend. In the 1980's, or even the 1990's, who would have been able to say how many people would be employed in a decade's time as web developers or graphic designers? That is not to say that a certain degree of re-skilling will not be necessary, but it is important to view such changes in their historical context – innovations in technology are always closely followed by innovations in the labour market soon after.

2.6 Indeed, though it must be stressed that this will only happen if it is used correctly, strong AI truly does have the potential to benefit *all* of society. If AI can be used to analyse and interpret health data and patient medical records, Doctor time can be more wisely spent on diagnosis and real patient care – benefitting us all. Strong AI could be used to more efficiently manage social housing, it could be used to control the usage of electricity, heating, water, it can create a shared transport system, to manage holiday planning, to better evaluate insurance claims. The list is, quite possibly, endless. AI is a neutral technology. Its benefits to society will be determined by how it is implemented.

3. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

3.1 One sector that we have chosen to highlight, and a sector that Braintree will be moving in to in the coming months and years head, is Health. Health and Artificial Intelligence has been an area of much discussion over recent months, and the health sector is often a sector where AI is met with the most scepticism. Indeed, it has been the subject of some considerable controversy of late, with Google DeepMind's work with the NHS allegedly flouting patient privacy laws.

3.2 Suspicion is understandable. In the age of data hacks and leaks, even if you take the NHS cyber-attack that we experienced earlier this year, people are reluctant to share their data. Data does not come much more personal than medical history. However, despite this, medicine is a sector which could benefit enormously from AI technologies. From its most basic application, simply think of the vast amount of data that the NHS has at its disposal. Think of the mountains of individual patient records and histories, transplant databases, even internal staff rotas. AI technologies can categorise, shape and manage this data in an infinitely more efficient method than human, still largely paper-based, systems. In a time where all public services are being asked to make efficiencies, we should embrace this possibility rather than shun it.

3.4 Away from general administration, AI's potential in the health sector is exponential. If we allow access to the right data, AI can be used to model diseases or tumours. It can predict their likelihood of growing or spreading, or it can be used to predict when they are most likely to occur. AI can find patterns in patient data to find correlations between certain symptoms or illnesses and try and interpret future ailments. All of this can greatly help patient outcomes, but fundamentally it helps doctors do their job better. Instead of being buried in administrative duties and routine medical analysis, they could concentrate more fully on patient care and higher level medical diagnosis. AI in this instance would not replace the role of doctors by any means, but it would free them to work where they are most needed.

3.5 AI also opens up the possibilities of accurate remote diagnosis. With AI, mobile applications could be created to, once you have plugged in your symptoms, assess all possible outcomes in combination with a patient's precise medical history to provide a diagnosis. With advances in mobile technology, it is not unreasonable to predict that this could soon include heart rate, temperature, and more. Anyone who has searched their symptoms online before may find this a daunting prospect. However, proper access to patient data would battle such inaccuracy. This would, again, dramatically free up doctor's time, keep potential patients from visiting health services if they were not sufficiently unwell, and ensure that those who are making appointments are those most in need.

3.6 This remote does not just have to be for diagnosis, either. It can also apply to patients who have long-term conditions, assisting them in when and how to take their medication, what dosage they need, repeat ordering their prescriptions when necessary or, with sufficient access to their data, helping them to manage their condition day to day.

4. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

4.1 Fundamentally it is up to the UK how it wants AI to develop, and how it wants to lead in this area. AI technologies will develop apace, with or without government involvement. Other country's governments will develop their own AI priorities and technologies and channel it as they see fit. Therefore, if this country wants to be at the forefront of responsible AI development, then it must lead from the front – AI won't wait for the UK to catch up. Government and industry must now rapidly work together to develop 'terms of engagement' to set the parameters of responsible and ethical AI innovation in this country.

4.2 The UK is a global brand in terms of rule of law, responsible corporate governance, and a benign state. If we can develop the first true framework in this country for leading AI developers to deliver advanced, safe, and responsible, AI solutions, they will (almost by default) have the trust of a global market. In

the same way other technologies developed in other parts of the world may not. This is a huge opportunity for UK plc to capitalise on, influence, and lead AI development at a global scale.

4.3 Government's role in AI development is however far wider than simply enabling innovation by creating the structures for the private sector to work within. As discussed above, AI thrives on 'big data'. Government is by its very nature the largest big data holder there is. Government therefore has a significant role/interest in the development of UK-based AI technologies, not only because the economy more widely can benefit from AI's positive growth, but also because the government can benefit directly. The government can benefit in multiple ways in terms of the effective running of public services and its own data management as well as securing the UK's data.

4.4 Security is a crucial part of the compelling reasons for the UK being at the forefront of the AI development. If the UK government takes an active role in the use of AI for its own data and systems it will therefore be actively participating in the development of securing its (and the public's) data. For if the government decides to step away from AI development, and leave government held data out of the equation, it will in fact become more vulnerable to nefariously developed AI designed to exploit/steal/damage its data. But if government develops AI solutions for the use of its data in collaboration with the industry experts (working within a framework as highlighted above), it will instead be working proactively to secure its data from malign forces.

4.5 AI's can be developed to help the government and its people make access to their data even more secure. For example, the public are rightly concerned (as has been recently demonstrated with NHS hacking) that malicious viruses can compromise our public services. In fact some question the development of AI as a potential route to developing even more viruses with malicious intent. But, though it may be true that some will develop AIs for nefarious reasons, that should not hold us back from developing AIs that make our data more secure.

4.6 As described above, strong AIs require significant amounts of data to be effective. Those developing AIs for malicious reasons will not have access to the same levels of data that those working with the cooperation of government will have. Therefore, those working on behalf of securing the data are naturally going to be in position to develop stronger AIs than their opponents.

4.7 Furthermore, organisations like our own can help in creating significant levels of security that opponents will not be able to emulate. To give a practical example, word/numerical based passwords have become an everyday part of our lives and our data security. However, they are not a very secure form of protection, and are a code that hackers are increasingly able to crack. Strong AIs, like the ones Braintree is developing, can expand the security options available to us at a significant rate – all of which can be used to protect our data.

Rather than a state employee simply using a password to access sensitive data, they could in theory provide a password, retina scan, fingerprint, breath test, temperature test, and so on. This would however not be the administrative burden that it may sound, using the right AI based sensors on the terminal(s) where the data is being accessed from, these readings could all be taken at once, autonomously, without bothering the employee. A malicious AI/virus would struggle to obtain enough data to emulate these different levels of security simultaneously.

4.8 As described in the example above, a Braintree solution identified a multimillion pound fraud at an international retailer because it could discern anomalous patterns in its big data (without even being asked to look for them specifically). The potential for AIs to interrogate our state level data and systems is significant not only because of AI's ability to establish potential efficiencies that can be exploited at policy level, but also the opportunity for AIs to establish security gaps and advise accordingly.

4.9 In essence, rather than the government taking an overly cautious approach to AI development, it should instead embrace and harness it so that the country's data and systems can be secured, and then maximise that data so that policy makers can make more effective and evidence-led policy decisions. Again, as stated above, if the government works quickly to establish an authoritative framework for the sector to develop in, it can have more confidence that those developing solutions for state level data are working to the wider benefit of society. In turn, the technologies that are developed through this ongoing process will be of significant economic value internationally.

Regulation

4.10 AI technology in and of itself does not necessarily need to be directly regulated. However, some of the sectors where it has potential to make a significant positive impact is of course in highly regulated areas, such as medical innovation, security, energy distribution, and so on. It is therefore important that the specific application is regulated as appropriate in the context of its use. AI is an extremely powerful tool, and, as with any powerful tool, it is necessary to create rules as to where, how and who can use these solutions.

4.11 It is therefore important that government is abreast of the increasing power of AI and the expanding role it may be about to play in its citizen's lives. It is important that the government enables innovation and shapes it positively. To hold it back, to try to rigorously control it, would be largely futile and counterproductive. Larger technological companies at the forefront are going to continue to develop AI at pace and scale, with or without the UK government's acquiescence. If the UK turns its back on developments here, they will simply go elsewhere. Possibly where state actors are less rooted in a culture of a rule of law, or democratic accountability, and therefore the AI innovations that emanate from these places may naturally be creatures of where they are born.

4.12 The UK must leverage its position as a home of responsible business practices, as a tech hub (silicon roundabout, etc.), its robust legal system, and significant access to finance, to make it the world leader in AI innovation. That way it will attract the best talent, the most responsible AI innovators, and the leading technologies. In turn it will reap the rewards and benefit from the early adoption of leading AI technologies that boost the UK economy and improve public services.

5 September 2017

Mr Philip Bree and Mr Jaafar Almusaad – Written evidence (AIC0039)

Mr Philip Bree and Mr Jaafar Almusaad – Written evidence (AIC0039)

[Submission to be found under Mr Jaafar Almusaad](#)

Bristows LLP – Written evidence (AIC0097)

Chris Holder
Partner
Bristows LLP
Head of Robotics and AI, Commercial Technology Department

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- 1.1 In order to answer this question, it is important to define and understand exactly what we are talking about when it comes to 'Artificial Intelligence' ("**AI**"). This term is being used to categorise a range of new technologies that are all inter-dependent and, in our view, is often being used incorrectly.
- 1.2 The Committee on Legal Affairs of the European Parliament produced a report containing recommendations to the Commission on the Civil Law rules on robotics on 27 January 2017 ("**the European Parliament Report**"), the House of Commons, Science and Technology Committee Report on robotics and artificial intelligence was produced in September 2016 ("**the House of Commons Report**"); and the RAS Robotics and Autonomous Systems Report by the Special Interest Group produced in July 2014 (the "**RAS-SIG Report**") are focussed on what are known as 'Robotic Autonomous Systems' ("**RAS**"), which contain within them the separate disciplines of robotics, big data, machine learning, autonomous systems, the Internet of Things and AI. It is our view, backed up by the references in the above reports, that AI is a subset of RAS and that this distinction needs to be made because of the confusion that may follow as a result.
- 1.3 As it is one of the main recommendations from the European Parliament Report that certain robots should be registered at a 'to be established' regulatory body, then the question has to be asked, what is a robot? If a robot is to be granted a certain legal status (which is another recommendation), then what exactly is it?
- 1.4 AI is not the same as a robot. AI is, in its purest sense, software that allows machines (robots/autonomous cars/computers (IBM Watson, Google Deepmind)) to sift through vast amounts of data in order to establish certain correlations or data points, which in turn provide hitherto unimagined insight into what may be happening in certain areas, for example, the interrogation of the human genome for scientific research purposes.

- 1.5 Autonomous vehicles (driverless cars) use AI to analyse vast amounts of data collected by the car's sensors as it drives down the road in order to recognise objects and road markings (amongst other things) so that the car can navigate around its surroundings – but to term this technology as being AI is to miss the point. AI is used within these RAS to “make decisions” based upon algorithms which have been programmed into computer systems. If the algorithms do not work properly, then that is a functionality issue and the software is at fault. The current level of 'AI' within a driverless vehicle does not allow a car to make its own decisions as a human being would – it is merely hardware and software working together, albeit in an extraordinarily complex way.
- 1.6 The current sophistication of AI is, therefore, not really that different from current technology especially when it comes to reviewing issues of liability for non-performance. If the software does not work, then the software developer is at fault because there is a functionality issue.
- 1.7 However, in the future, when machines start to interact with other machines without any human involvement and actually begin to perform functions based upon these interactions, then traditional legal, societal, ethical and technological frameworks become eroded. All of these have been developed to date with the human being as the central actor on this stage – companies, partnerships and other legal constructs aside.
- 1.8 The minute that RAS moves away from this human centric construct is the minute that things change completely. As RAS become capable of making their own decisions, enter into contracts, create software and hardware and generally act as humans traditionally have, this creates issues for traditional constructs like law and this is why the importance of definitions comes to the fore.
- 1.9 Regarding the factors that may accelerate or hinder this development, these may be described as being largely two sides of each particular issue. For example, public trust in the autonomous vehicle industry will either exist (and therefore help to accelerate the industry's development) or it will not. If it exists, this will help the technology develop. If it does not exist, then future development is at risk.
- 1.10 Public trust is one example of this, others would include the following:
- Well informed regulation/legislation;
 - Continued advances in technology (battery technology, computing, engineering etc);
 - Education and creation of a skilled workforce;
 - Continued investment in research and development; and
 - Reduction in cost of usage.

Is the current level of excitement which surrounds artificial intelligence warranted?

- 2.1 As outlined above, AI is part of a larger, more complex and more integrated set of technologies which have been loosely defined as representing 'Industry 4.0'.
- 2.2 These technologies working together, will, we believe, have a profound impact over the population of this planet and so we believe that this current level of excitement is warranted.

How can the general public best be prepared for more widespread use of artificial intelligence?

- 3.1 Increasing public awareness of what AI and RAS are and how they are likely to be used, will be helpful in managing the public's expectations, reversing our inherent resistance to change and help to reduce our fear of the unknown.
- 3.2 For example, in dealing with the fear of potential, large scale job losses, it should be made clear that human development has been characterised by much technological change over the centuries and much like the automobile was seen as a threat to jobs and the 'way of life' at the beginning of the 20th century, such fears were shown to have been misplaced given that a new industry was created, alongside millions of jobs. The same should be true of developments in the application of AI and RAS.
- 3.3 The easiest way for the public to get used to and prepare for the advances in these new technologies is to see them in action, to use machines that are designed as RAS, to be educated about the benefits and not to be continually bombarded by horror stories or what may happen 'when the robots take over'.
- 3.4 Much like genetic engineering, chemical engineering, atomic research and development and countless other technologies, there is always a 'bad' side to what can be developed but there are also the tremendous benefits to humanity that should be taken into account – and Government has a large part to play in reinforcing this and regulating against any harmful side effects.
- 3.5 STEM subjects taught at schools will enable more people to understand what is being talked about, as well as training the necessary workforce to develop and manufacture these machines and supporting systems.
- 3.6 Data collection, protection, analysis and storage are extremely important issues to be addressed and understood. Current attempts to try and create 'trusted' parties who will, on the one hand, gather huge amounts of data and,

on the other hand, use such data to create new products and services, should be examined closely.

- 3.7 Intellectual property rights and commercial imperatives should continue to play a role in framing how datasets are made available to public and private entities and the value in such rights should not be eroded by developing additional regulatory bodies to oversee these areas.

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

- 4.1 Society as a whole stands to benefit from the development and use of AI and RAS, provided that the technology is developed and implemented responsibly.
- 4.2 At the moment, it is the technologically savvy that are benefitting the most as new businesses spring up (some of which grow very quickly indeed) and existing technology companies deploy AI and RAS in their products and services. This pace of company development and concentration of wealth into the technology classes is expected to increase as RAS becomes more widely used – but the potential benefits to ordinary people should not be underestimated.
- 4.3 Healthcare solutions using robotic carers, smart cities, smart homes, driverless cars, space and sea exploration, reduced global emissions and many more areas for RAS application will provide wide ranging benefits – it is up to technologists and Government to plan ahead in order that these technologies can be used for good and jobs can be created in the UK on the back of them.
- 4.4 What could be problematic in the future is vast wealth concentrated in smaller and smaller businesses that employ fewer and fewer humans but run vast estates of RAS. The implications for taxation, employment and public services are self-evident. Therefore, debates around the creation of a ‘robot tax’ or a ‘universal basic income’ to everyone should be held to better prepare society for any potential implications in the future.

Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

- 5.1 Public trust is crucial to the successful propagation of RAS across society. Further, in order for the public to readily adopt the RAS technologies, clarity of accountability, liability and routes of recourse and redress are necessary. If something goes wrong, people generally want to lay blame at someone else’s door. As a result, systems must be put in place for when incidents occur and

things go wrong, and the public should be aware of these routes of redress and compensation.

- 5.2 Trust is built through understanding, which is achieved by effective communication of the technology's potentially substantial benefits for society, along with the negative effects and how they will be dealt with. Engagement with the public, in a way that demonstrates public opinion is being actively considered and incorporated with regard to its implementation, will be extremely important.
- 5.3 If the public believes that RAS will be both beneficial and safe, the development and up-take of the new technologies will be greater.
- 5.4 One way to effect this trust is by creating industry standards to guarantee what is safe, as judged by a regulatory body trusted to monitor and certify such standards against a code of practice for the use of RAS.
- 5.5 It has long been considered that public trust in new technologies is directly affected by the amount of regulation that is put in place and so industries such as the aviation industry are often cited as examples where robust regulation increases public trust in an otherwise inherently risky process.
- 5.6 Media outlets should be encouraged to view the positive aspects of RAS rather than the negative ones.
- 5.7 More funding needs to be made available to research bodies and universities, particularly in the light of Brexit and the extremely high percentage of new businesses that are funded by the EU.
- 5.8 Primary and secondary education should include computer science and RAS on the syllabus as well as the STEM subjects – and more emphasis should be put on the participation of woman in these areas.
- 5.9 More than anything, a coordinated Government approach is required across all Departments. These new technologies are currently recognised as being of particular importance to the UK economy for the future and they should not be forgotten or hampered by Brexit or a disparate, uncoordinated approach across Government.

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

- 6.1 It is our view that virtually all industry sectors will benefit from RAS, much like existing technology and the use of the Internet has changed the way we live our lives across all areas.

How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- 7.1 Data is currently the steam that drives the pistons of technological evolution in our society and therefore its importance to future technological innovation cannot be underestimated.
- 7.2 Collecting 'clean' and untarnished data is expensive and time consuming. Some companies profit almost solely from such processing, which means such companies will wish to protect their industrial know-how and trade secrets and will be reluctant to make it available, free of charge, for general use.
- 7.3 Despite the ideal of a global community where RAS are free to use information and interact across jurisdictions free of charge, it seems more likely that a rights-driven, hard-nosed and economically focused commercial approach will occur.
- 7.4 Perhaps it is time to look at ways in which data can be made available outside of such approaches and so perhaps the models of usage applied to existing technology, like open source software, should be examined. In relation to data held in the public realm, it should be understood that such data is potentially very valuable, and any public – private commercial arrangements to exploit such data should be designed to extract as much value for the public as possible.

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

- 8.1 The key ethical issues that arise from the ever-increasing pervasiveness of AI and RAS in society need to be addressed, particularly in relation to: i) safety and control; ii) privacy and consent; and iii) diversity.
- 8.2 Safety and Control – it will be vital for public trust that RAS are safe and can be controlled if required. There are many ethical, moral and legal issues if RAS are not inherently safe or controllable.
- 8.3 Privacy and consent – RAS may be used in homes and in public spaces and therefore people will not necessarily know that they are present and operating. The collection, storage and analysis of data in these circumstances needs to be highly regulated if public trust in RAS is to be established. We will interact with RAS in ways hitherto not envisaged and the anthropomorphism of robots will lead to people looking at machines in a completely different way than is the case at present. All of which needs to be taken into account when looking at laws that deal with data collection, storage and analysis.

- 8.4 Diversity – RAS runs on data and data needs to cover all aspects of our society, not just a rich, white, middle aged and male section of it. Decisions will be made by RAS based on such data and so careful analysis of the datasets being used will be necessary.
- 8.5 What is it to be human? – Such questions will need to be addressed once the use of brain-computer interface technologies, exoskeletons, robotics limbs and organs become prevalent. If people become more machine than human, does this impact upon their rights in society?
- 8.6 Ethics in general – will play a central role in framing the difficult questions and how society needs to deal with them. The outcomes of such debates will heavily influence the law makers and regulators of the future.

In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable? When should it not be permissible?

- 9.1 In order to maintain security and privacy, the ‘black boxing’ of AI may be necessary – but only to the general public.
- 9.2 In order to maintain public trust and in order to prove that AI and RAS are inherently safe and controllable, transparency will always be required to some degree via specific laws and regulations.
- 9.3 The new technologies may be used for bad as well as good and so certain aspects of it may need to be made subject to the same sort of regulatory framework as is currently applied to nuclear proliferation and other weapons grade technologies available today.

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- 10.1 Government should look at regulation specifically as it relates to each application of RAS, across industry sectors.
- 10.2 For example, regulation to enable the testing and development of autonomous vehicles will be different to that required for the use of care robots in a person’s home.
- 10.3 A ‘robot law’ seeking to regulate RAS across the board, as was envisaged by the European Parliament Report, is not necessary in the UK, we believe, because the UK has a common law system. This system has

proved to be very successful in the past in applying existing common law principles to new and emerging technologies.

- 10.4 A nascent industry does not need more product liability laws. These may kill off development and since we already have such well-developed concepts like 'negligence' and 'duty of care' at common law, these add a flexibility of approach that will benefit the future development of new technologies.
- 10.5 When the time arises that RAS communicate with each other and start making decisions, entering into agreements or creating works, then existing laws may need to be adapted, for example in relation to Contract Law and Intellectual Property Law. At present, both these areas of law require the presence of humans to either form contracts or create and register patents, for example. As the technologies develop, we would suggest that specialist technology lawyers need to be consulted to advise on the best ways to approach the regulation of the same.

What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence

- 11.1 In Europe, the new General Data Protection Regulation will provide a new framework for the collection and storage of data. How such a regulation has been developed to cater for changes in technology is a good example of how difficult it is for law to keep pace with technology and lessons should be learnt.
- 11.2 The European Parliament Report provides a series of recommendations to the European Commission. This approach was largely adopted by the House of Commons Report and is a very helpful start in highlighting the many issues that RAS creates and how they may be dealt with.
- 11.3 The USA continues to be a leader in the field of RAS and it has many commentators who write extensively about the topic. Their views and experiences should be harnessed in order to provide insights into the industry.
- 11.4 In Europe, aside from the work being done by the European Parliament, the European Robotics Forum is an excellent example of an organisation that represents the interest of industrialists and academics operating in the RAS industry and its views should be actively sought in developing legal frameworks and approaches.

11.5 Finally, from a legal perspective, the IBA (the International Bar Association) which has a dedicated Technology Law Committee, should be canvassed for its views on the international implications of RAS and AI.

5 September 2017

The British Academy – Written evidence (AIC0213)

Submission to the House of Lords Select Committee on Artificial Intelligence

Introduction

1) The British Academy is the UK's national academy for the humanities and social sciences. A Fellowship of over 1200 of the country's leading academics, the Academy received its Royal Charter in 1902. It exists to promote and champion its disciplines, and awards funding to researchers at all career levels. This submission represents the views of the British Academy, and not one specific individual.

2) The humanities and social sciences provide a critical lens through which Government and society can address the wide-ranging challenges we face today. From security to health, climate and demographic change and technology, the humanities and social sciences can provide a crucial means of focussing on the issues facing our world, and offer solutions to seemingly intractable problems.

3) The British Academy has a strong track record in bringing the expertise of our disciplines to bear on the economic, ethical and social issues of emerging technologies. In 2012, we co-hosted a workshop with the three other national academies on human enhancement and the future of work¹⁹⁰ and have hosted a series of seminars on Robotics, artificial intelligence and society.¹⁹¹ 4) We have also been pleased to work recently with the Royal Society on 'Data Management and use: Governance in the 21st century'¹⁹². This multi-disciplinary project was co-chaired by Professor Dame Ottoline Leyser FRS and Professor Genevra Richardson FBA, with members of the working group drawn from a range of disciplines.

5) It is from the work of this project that most of our submission is drawn, whilst the project covers all aspects of data management and its use, this clearly includes the management and use of data which feeds, and is generated by, artificial intelligence. The definition we have used for artificial intelligence in this submission is the same as that in our 'Data Management and use: Governance in the 21st century' report: "an umbrella term for the science of making machines smart"¹⁹³.

¹⁹⁰ https://www.britac.ac.uk/sites/default/files/12308_academy_HE_report_2012_web.pdf

¹⁹¹ <https://www.britac.ac.uk/sites/default/files/Robotics%20AI%20and%20society.pdf>

¹⁹² 'Data management and use: Governance in the 21st century' British Academy and Royal Society, June 2017

<https://www.britac.ac.uk/sites/default/files/Data%20management%20and%20use%20-%20Governance%20in%20the%2021st%20century.pdf>

¹⁹³ 'Data Management and use: Governance in the 21st century', Glossary. P.94.

Questions:

The pace of change

6) The great economist and historian of technological change, Chris Freeman, identified five waves of technological change: the first was the industrial revolution, particularly around textiles, in the late 18th century; the second saw the age of steam and railways from the mid-19th century; the third was the age of steel and electricity during the later 19th century; then came the age of oil, cars and mass production and more recently the age of information and communication in the later 20th.

7) We are now in the midst of further rapid change across a number of interwoven dimensions: notably in artificial intelligence, in biotechnology, and in materials. This sixth wave is moving at lightning pace. The multiple and mutually-reinforcing dimensions we are witnessing present spectacular opportunities. But, like the waves that preceded them, they also entail fundamental dislocation and the redefinition of work and of other activities. In this case, the pace and nature of change appear unprecedented, as do the opportunities and potential dislocation.

Impact on society

8) Artificial intelligence will impact on society in a variety of ways, both positive and negative, some of which it is not yet possible to predict. Our series of lectures examined the implications, from a variety of perspectives, with one participant quoting William Gibson – “The future is already here, it’s just not very evenly distributed”. Professor Christian List FBA also highlighted the need to mitigate the problems of winners and losers. “As a society, we need to think hard about more egalitarian schemes that distribute the fruits of productivity growth fairly, which we can achieve through automation.”¹⁹⁴

9) Artificial intelligence has huge potential to contribute, but the ethical implications of these choices need to be considered. For example, would residents in care homes be happy interacting with a robot carer, rather than a human being? Additionally, if robotics is going to make hundreds of thousands of job roles redundant do we have a duty as a society to consider whether or not we want to allow this to happen? Is it better to stand in the way of such technological developments for the ‘greater good’ or is it preferable to allow these systems to develop, and potentially address negative consequences after the fact?

¹⁹⁴ <https://www.britac.ac.uk/sites/default/files/Robotics%20AI%20and%20society.pdf> British Academy event ‘Does AI pose a threat to society?’ 1st March 2017, University of the West of England.

10) Artificial intelligence will undoubtedly impact on the skills individuals will need to thrive in the labour markets of the future, and during the likely period of disruption. Increased automation is predicted to lead to more demand for high-skilled jobs, particularly in major occupation categories 1-3 in the Standard Occupational Classification (Managers, directors, senior officials, professional occupations and associate professional and technical occupations)¹⁹⁵. Computers are unlikely to be able to perform general behavioural and non-cognitive 'soft' skills necessary for collaboration, innovation, and problem solving such as resourcefulness, creativity, abstract reasoning, and emotional intelligence.

11) Demand is growing for individuals to be equipped with these higher-level skills which they can deploy in different contexts, whether in a career which may cross many sectors of employment or within a research community which is increasingly interdisciplinary. It is essential that our workforce is equipped with the skills to allow them to cope, adapt and thrive. In a world that will be far more complex and interconnected the arts, humanities and social sciences are ideally placed to deliver these skills – resilience, adaptability, flexibility, adapting to change, navigating uncertainty are some of the core skills provided by these disciplines.

12) In a context of rapidly rising computing power, the ability to reason and manage in a data-driven environment will certainly become a crucial skill for the next generation of graduates and workers as demonstrated by the British Academy policy report, **Count Us In**¹⁹⁶.

13) More work is needed, from economists, computer scientists and others to understand the potential changes to the labour market. Government needs to be looking at the jobs likely to be in existence in 30 – 50 years' time, and plan accordingly to invest in skills in schools as well as lifelong learning opportunities to help improve the resilience of the labour force. This could most usefully be done in the context of the developing industrial strategy; the industrial strategy should not develop without a significant focus on the challenges posed to the economy and labour market by artificial intelligence.

14) The recent report by Professor Sir John Bell on the life sciences emphasizes the important potential offered to the NHS by data, and its ability to create advances in imaging and pathology via artificial intelligence. There are a myriad of issues here which the humanities and social science disciplines can play a valuable role in addressing.

15) We would support ongoing reviews in to skills levels and specific jobs and sectors likely to be affected, negatively and positively, by the growth of artificial

¹⁹⁵ The Economist Intelligence Unit (2015). *'Automated, creative and dispersed: the future of work in the 21st century'*

¹⁹⁶ *'Count Us In: Quantitative Skills for a New Generation'*

https://www.britac.ac.uk/sites/default/files/Count-Us-In-Full-Report_0.pdf published by the British Academy, June 2015

intelligence. Schools should be preparing all pupils for a working life in which multiple jobs and careers are possible, and should ensure all pupils are fully computer literate.

Public perception

16) We would support moves to improve levels of public literacy on artificial intelligence, but this needs to be done in an informed and balanced manner that avoids unnecessarily fueling levels of public anxiety and any sort of public backlash which might inhibit the development and use of artificial intelligence. Fundamental to the success of improving the public's understanding and engagement with artificial intelligence is the need for trustworthy governance structures surrounding them and for the public to have trust in them.

Ethics

17) There are considerable ethical questions and dilemmas posed by the development of artificial intelligence, particularly with regard to robotics, and the likely future growth of robots in roles formerly carried out by human beings. Particularly challenging is the idea of devolving responsibility to robots. "What makes someone or something a moral agent?" When an artificial intelligence system causes harm, who should be held responsible? Maybe whoever designed the software? Should it be the manufacturer? Should it be the operator? It will be an important job for philosophers, computer scientists, lawyers and society at large to think about how to refine our moral codes.¹⁹⁷

18) The British Academy recognises that many of the choices that society will need to make as data-enabled technologies become more widely adopted can be thought of as a series of pervasive tensions, which illustrate the kinds of dilemmas that society will need to navigate.

19) Many of the tensions alluded to in the questions raised by the committee are explored in more detail in the British Academy and Royal Society report, in relation to data management and its uses¹⁹⁸ (Box 4, p41). The nature of these tensions are such that they resist linear, ad-hoc policy solutions.

Government

¹⁹⁷ <https://www.britac.ac.uk/sites/default/files/Robotics%20AI%20and%20society.pdf> British Academy event 'Does AI pose a threat to society?' 1st March 2017, University of the West of England.

¹⁹⁸ 'Data management and use: Governance in the 21st century'. P.41 "Box 4. Framework for social and ethical tensions"

20) Government can play a significant role in ensuring that there is an effective framework for data governance. The BA and RS report concluded that whilst existing frameworks provided much of what is sufficient for today there is need to develop a new framework to cope with the challenges of the future. Whilst our project covers all aspects of data management and its use, this clearly includes the management and use of data which feeds, and is generated by, artificial intelligence.

21) Firstly, there is a need to develop a set of high-level principles for the management of data and its use.¹⁹⁹ The overarching principle is that systems that govern data need to **promote human flourishing**. This principle is intended to provide an orientating mission that has 'the human' at its centre. At moments of contention, it should serve to reflect the fundamental tenet that society does not serve data, but that data should be used to serve human communities.

22) The concept of 'human flourishing' is deliberately broad. It emphasises the nature of human wellbeing and recognises the importance of context and the role of competing interests and values. Arguably this overarching principle could be usefully applied to the use of artificial intelligence – artificial intelligence needs to serve society, not society serving artificial intelligence.

The four additional principles are intended to inform and shape all aspects of data governance. They are that all systems of data governance:

- protect individual and collective rights and interests
- ensure that trade-offs affected by data management and data use are made transparently, accountably and inclusively
- seek out good practices and learn from success and failure
- enhance existing democratic governance.

23) The report also identified a clear need for **a new body to steward the data governance landscape as a whole**. We expect that a stewardship body would primarily recommend actions to others, but it may also need the capacity to carry out some functions itself if they could not be performed elsewhere, being careful to not duplicate existing efforts. It would be expected to conduct inclusive dialogue and expert investigation into novel questions and issues, and to enable new forms of anticipation about the future consequences of today's decisions.

24) Government also clearly has a role in play, primarily through the development of the **industrial strategy**, to plan effectively for the impact that artificial intelligence will have on future jobs and skills. It is vital that the public are equipped with the skills needed to adapt to a changing labour market, and

¹⁹⁹ 'Data management and use: Governance in the 21st century'. P.51

that the education process equips people for the impact automation may have on jobs, and their wider lives, in the coming decades.

25) The UK has a competitive advantage in this field, and has the opportunity to be a world leader in this sphere. To capitalise on the advantages that can be derived from artificial intelligence the Government needs to continually take account of the wider social and ethical issues thrown up as the technology and its applications develop. This will require a multidisciplinary approach, looking not just at technology and its applications, but at societal impact, forecasting and modelling, ethical implications and the views of the public. The British Academy is well placed to continue to contribute to this vital work with the other national academies.

Barbara Limon, Helen Gibson
British Academy

11 September 2017

The British Institute of Facilities Management (BIFM) – Written evidence (AIC0205)

BIFM submission to Call for Evidence

1. The British Institute of Facilities Management (BIFM) welcomes this opportunity to provide feedback and evidence on the state and impact of artificial intelligence (AI).

About BIFM

2. The BIFM is the professional body for facilities management (FM). Founded in 1993, we promote excellence in facilities management for the benefit of practitioners, the economy and society.

3. We represent and support over 17,000 members around the world, both individual FM professionals and organisations, and thousands more. We do this through a suite of membership, qualifications, training and networking services designed to support facilities management practitioners in performing to the best of their ability.

4. We also provide guidance and support research that will help increase workplace productivity which will ultimately contribute to raising standards, a happy workforce and healthy economy, and provide a platform for meaningful and evidenced debate on issues of importance.

5. Based in the UK, BIFM's global reach has been formalised during the last few years by establishing regional operations in Ireland, the United Arab Emirates and Nigeria. In total, BIFM is represented in 80 countries across the world.

About the BIFM Technology Research Task Group- AI Sub-Group

6. BIFM often works together with industry experts to ensure that it can provide a voice and expert knowledge where needed on technological matters. This submission was prepared with the expert input of the BIFM Technology Research Task Group- AI Sub-Group. This group advises, collaborates and works with BIFM representatives to share knowledge in technology advancements with BIFM, its members and across the FM profession.

7. In addition to being FM experts, these BIFM members work with AI in smart buildings and intelligent management systems to help manage businesses' estates. The submission should be read against this background.

About FM

8. "Facilities management is the organisational function which integrates people, place and process within the built environment with the purpose of improving the

quality of life of people and the productivity of the core business”.²⁰⁰

9. Facilities management encompasses multi-disciplinary activities within the built environment and the management of their impact upon people and the workplace. FM contributes to the everyday functioning of hospitals, airports, universities, down to ordinary businesses. By making the workplace as efficient as possible, the facilities manager has a major role to play in making the UK a more productive place²⁰¹. At the same time, without FM support, the economy would grind to a halt.

10. The health of the wider FM industry, which accounts for around 7% of GDP²⁰², has a major impact on the overall UK economy and plays a positive role in supporting the government’s climate change targets and societal and modern slavery programmes amongst others.

11. Key facts about the FM Industry:

- The UK FM Industry accounts for around 7% of the UK’s GDP²⁰³
- The value of the FM sector is put at up to £120 billion²⁰⁴
- FM employs almost 10% of the UK’s workforce²⁰⁵
- In parts of the industry, up to 24% of the FM workforce are EU nationals²⁰⁶
- An effective workplace can improve productivity by 1-3.5%, potentially delivering a £20 billion uplift to the UK economy²⁰⁷

Please find below a response to your specific artificial intelligence (AI) inquiry questions:

The pace of technological change

Q1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

12. One could argue that AI is at an embryonic stage, the outer edges of its shape of design have been drawn providing a framework for advanced function and deeper learning. At the same time, the current state of AI is already at a

²⁰⁰ [International Standards Organisation ratified definition](#)

^{201, 8} [The Stoddart Review – The Workplace Advantage](#), (December 2016), 42p.

^{202,4} FM Business Confidence Monitor, (May 2015), 12p.

⁵ Value Judgement, Facilitate, FM World, May 2017, p. 49

^{6,7} Has Brexit hit home yet? Insights into facilities management, Issue 17, p. 17-18

significantly developed state- the progression of all aspects of computing as well as economic drivers have allowed a steady progress from advanced calculations to problem solving and now modelling the human brain.

13. The pace of development is only to accelerate in the coming 5-20 years owing to the consistent advances and dependency on technology in our work and personal lives. The future shows even more advanced applications in nano-technology as well as quantum computing, so with evolving frameworks along with Moore's law (the number of transistors per square inch on integrated circuits has doubled every year since their invention, allowing for increasing processing power and computation speed, a trend which will continue into the foreseeable future) and further societal adoption the development will continue to integrate with more and more granularity. Automated transportation, cyborg technology, connected work/home spaces and supply chain logistics, replacing workers of dangerous jobs and care assistants are key areas that AI will help advance over the next decade.

14. Due to the potential for such far reaching change across society the main factors on how it will unfold, and the pace it will unfold, will depend on the balance that can be struck as society adapts to functioning and existing in a connected world.

15. Access to large amounts of big data, faster equipment and processing speed and advanced machine learning are technical factors contributing to the advancement of AI. Technical factors and societal factors (i.e. the public's level of trust in a safe application) are the two factors likely to either hinder or accelerate the speed of AI growth and its adoption.

16. In addition to the above, it is important to stress that the commercial benefits of AI for FM (and beyond), mainly automation and improved customer service, are enormous and the potential for true progress exists at the heart of AI.

Q2. Is the current level of excitement which surrounds artificial intelligence warranted?

17. Every hour millions of devices, the Internet of Things (IOT) and other platforms gather billions of data points. The rate of accumulation is growing exponentially and we are now able to leverage the insight and learning this data can offer but the only way we can leverage the power and value of such data is through computing power and artificial intelligence. Advancements in machine learning, reasoning and perception are helping us enhance people's quality of life in areas including education, transportation, healthcare and building automation and management. The potential to uncover new insights, solutions and predictions from this dataset is very exciting.

18. The promise from AI is an enhanced experience alongside automating routine processes. If it delivers on this promise it will open further doors and harness a potential which does warrant the excitement.

19. However, this will be dependent on effective human input, in terms of control and communication, to ensure AI does not become a factor that has a negative impact upon our lives both in a work and personal sense. In addition to this, we should ensure that we do not become dependent on AI or allow it to become a dominant tool for facilitation across a range of areas.

Impact on society

Q3. How can the general public best be prepared for more widespread use of artificial intelligence?

20. Advancement in AI (machine learning, reasoning, perception) is helping us enhance people's quality of life in areas including education, transportation, healthcare, manufacturing, building automation and management. The potential to uncover new insights, solutions and predictions from the further use of data is very exciting. However, trust is hard to earn and easy to lose. Those companies looking to leverage AI's value must reassure the public as to the security and privacy of this data. Facebook, Amazon, Google, Apple, Microsoft are but some of the largest companies currently using AI. These companies are in effect rebranding AI to their products such as SIRI, Facebook Photo Tags and Google Ads. The public are being introduced to AI via these products and platforms. Softening the AI term and transforming it into usable consumer products is introducing the wider public to the power of AI although they do not know it. Business should be more open about this, and educate its consumers about the potential and its limitations so the public can understand AI better.

21. With the potential around AI being so large, its impact will be everywhere. The capability to revolutionise entire industries over very short periods of time, if not properly administered, will lead to widespread issues at every level. Strong leadership is needed that both understands and can harness AI's potential. The policing of this space is also a unique challenge within an ever-changing environment to anticipate and avoid potential problems. It also requires clear and fair governance to make sure this learning system does not cross certain lines.

22. It is important to remember that technology should remain an enabler for progress. Indeed, the widespread use of technology in the past has been an effective enabler allowing us to advance academically, technically and functionally in developing our lives across service, financial and educational forums. However, what is the tipping point at which society is no longer happy for AI to perform its function without supervision or evaluation? Which is the

point where we do not want to be dominated by an AI application, which was created in relation to socio-political necessities? At which point do we become concerned about privacy and overall ownership of information, and specifically data? To use examples, the use of Google, Facebook and other tech services have accelerated to such a point where it has become difficult to know, individually or collectively, how much we are in control of information and what are the potential impacts (positive and/or negative) arising from this.

23. For these reasons, Government and commercial/large scale users have a duty of care to ensure two things. Firstly, the general public should be educated about AI, its potentials and the changes it will bring about. Secondly, that society continues to reap the benefits of this technological progress. Government, together with key stakeholders, should have an ongoing dialogue which addresses these two points.

24. In addition, education and lifelong learning have a key role to play in mitigating the changes that will be occurring in the workplace, where AI will change the way people work. Young people must be familiar with the technology when they enter the workplace. To help prepare those people whose jobs will be most affected, programmes of lifelong learning should be advocated to retrain people to ensure they retain marketable and needed skills.

Q4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

25. Big business and government/state will benefit the most from the development and use of artificial intelligence and data unless the journey in the roll out of AI is so disruptive that the current model is changed.

26. Business will be benefitting especially because of the ability to utilise AI in furthering their profit and growth. From Facebook image recognition using 120m parameters to automatically tag users, Siri, Cortana, Alexa voice assistants processing our voice instructions, and Amazon tailoring its purchase recommendations for us.

27. In addition, those countries with large gross domestic product (GDP) and where the government is communicating and supporting effective integration of AI, in both private and public companies, will be mainly benefitting too.

28. In comparison, generally those with limited or no access to AI both in business and personal arenas will be gaining the least. At the same time, people doing roles that can be automated, like lorry drivers to call attendants, will lose out along with areas of society that struggle to integrate in a connected world.

Up to 30% of UK jobs could be impacted by robots utilising AI²⁰⁸. This can be mitigated by integrating AI universally within society where possible. At the same time, in those businesses where workers will lose out, new jobs will be created because of the use of AI. Lifelong learning is essential here to retrain and retool people, even though this will require investment and training opportunities. Proactive societal diversification will be needed to allow AI to be fully embraced through its own journey.

Public perception

Q5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

- Question taken together with question 10- **The role of the Government**

Q10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

29. There is an element of fear when it comes to AI and understandably so. Workforce replacement and technology collecting data about our everyday movements and actions reflect negatively on AI.

Previous revolutions of technological change have not led to an overall loss of jobs, but rather have disrupted the types of jobs people do and the way we work. For example, the retooling of the old textile factories led to job losses for workers, but new jobs were created.

30. Likewise, AI will create jobs as companies look to leverage it as a tool to help minimise operating costs and explore new market opportunities. There needs to be transparency and reassurance given to the people that AI has more potential for good.

31. Government, public bodies and private companies should begin to lead the dialogue showing the benefits and talking about the risks to help stimulate the conversation. Government in first instance should learn from the people creating the technology set and business leaders in the space. Public figures and specialists in this space are beginning to share their thoughts with large audiences and even beginning to lead the debate – more of this is required.

32. Due to the far reach of AI a basic education and understanding should be had by all the public, to ensure a positive response and progressive growth of both AI

²⁰⁸ PwC, *Up to 30% of existing UK jobs could be impacted by automation by early 2030s, but this should be offset by job gains elsewhere in economy*, <https://www.pwc.co.uk/press-room/press-releases/Up-to-30-percent-of-existing-UK-jobs-could-be-impacted-by-automation-by-early-2030s-but-this-should-be-offset-by-job-gains-elsewhere-in-economy.html>, accessed 8th September 2017

and society as two connected elements. Whilst AI is not 'organic', effort should be made to integrate it within society for the public rather than private companies or individuals simply flooding the markets of business and education with AI.

33. Government has really three roles to play. Firstly, it should be a custodian of this dialogue, which should enable greater understanding and then adoption/acceptance. Government has got a series of tools available to enable the dialogue, through marketing campaigns (like for example Start4Life) to educate people, providing a platform for discussion such as a review or leaders group. In addition, Government can build on the technology curriculum available in schools to ensure that next generations leaving school understand the benefits and limitations.

34. Any such dialogue should ensure that it has got a variety of stakeholders engaged beyond big business, including but not limited to academics, educational bodies, professional bodies and consumer bodies as they have a big role to play in pushing out and amplifying the message. Life-long learning, as mentioned earlier, has got an important role to play, not just in making people aware of the changes and opportunities of the technology, but also to enable them to remain in an ever-changing job market by ensuring they have the right, required skills.

35. Secondly, Government should build on the dialogue and rise to the challenge to provide a wider framework strategy for AI in which to provide security and stability for a space which is difficult to control. There needs to be a framework strategy which guides the roll-out without stifling AI's growth. The overarching principle governing such roll-out is to protect the disadvantaged areas to make sure the progress does not create a tidal wave, or in other words that the rewards are reaped for the benefit of society and outweigh the disadvantages. Other key starting principles should be around security, accessibility and privacy. Once established, reform can be looked at to ensure the framework remains relevant and up to date.

36. At this point in time, strict governance enforced from Government downward will only limit the potential of these people and the technology. However, this does not mean that there should not be some regulatory tools in place to penalise businesses or individuals for example misappropriating the use of AI or acting outside regulatory controls in place. One such example of a regulatory instrument in place is the UK Data Protection Act which is in the process of being updated in line with the General Data Protection Regulations where consent is put at the heart of any data use.

37. Thirdly, Government should also play a role through leading by example and enable the adoption of AI within society by both encouraging research and innovation in this area, but also by encouraging adoption of the applications

within its own Government operations. From an FM perspective, that is the adoption of data driven decision making when it comes to the assessment, utilisation, analysis and resultant actions associated with the use of space (and therefore the highest element of the occupancy cost) and the workplace environment with its direct impact on productivity²⁰⁹ as but two examples. The application of tools such as Building Information Modelling (BIM), when designing new buildings. The forthcoming Industry Strategy should ensure that it includes such considerations from an FM perspective given that it represents c80% of whole life costs associated with the management and operation of the Built Environment.

Industry

Q6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

38. Most, if not all sectors have the potential to improve dramatically via the application of Deep Learning (a sub method of Machine Learning) and AI; healthcare, manufacturing (and any labour heavy unskilled sectors), transportation, customer services, finance, entertainment and even sport.

39. FM as a sector will continue to benefit from AI by way of further improving overall business intelligence and customer experience across all the sectors it supports in their core business activities.

40. Process orientated and predictable services will gain from a manufacturing efficiency perspective but from a job perspective they are likely to lose out.

Q7. How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

41. Data-based monopolies of some large corporations and the economies associated with them will need strong leadership and internal governance to prevent a continual repeat of the 'winner takes all' model. Alternatively, business need to have corporate governance safeguards in place that are part of its own fabric and as such would break the current model from within.

42. The use of AI by large corporations could potentially increase the likelihood of globalisation through use of a new forum. When it comes to data management, there should be clear and effective steps taken to distinguish and safeguard the rights of individuals within society in relation to the economy when observing the impact of AI. The use of the media has been something that large

²⁰⁹ The Stoddart Review – The Workplace Advantage, (December 2016), 42p.

corporations have benefited from previously with recent incidents where financial figures were hidden from the general public. To have a functioning economy transparency is key, where applicable, as is discretion. The difficulty is to keep the balance between those.

43. In terms of management and safeguarding of data there should be effective barriers where AI cannot be hacked or used in general to obtain sensitive data of any nature. This is where a Government-led dialogue is key, to ensure that there are some parameters in place wherein AI can operate. This will also ensure that there is transparency about obligations and restrictions for the benefit of society which in turn will help with societal acceptance of AI.

Q8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

44. The ethical implications of AI are its impact on society in terms of privacy, consent and safety. In addition, it affects diversity and has the potential to shaping democracy. AI is primarily a decision-based cause and effect mechanism. Human accountability for machine led decision-making would help to resolve negative implications by way of keeping the use of AI within acceptable and agreed parameters.

45. In terms of capturing and using personal data, the General Data Protection Regulation and the forthcoming updated UK Data Protection Act are a good start as they put the concept of consent at the heart of data use. They impose an opt-in principle which should help raise awareness of which data and why data is being captured. Indeed, people have got to be made aware of how data is being captured, who owns it and how it is /may be shared. AI relies on already existing datasets which are already being collected from people.

46. The development of AI has the potential to negatively affect unemployment rates (even now that they are at their lowest for some time). Those engaged with AI need to ensure that the benefits of AI outweigh its negative impacts. As mentioned before, a structured and well-defined roll-out of the technology is needed, as well as encouraging people to retrain and diversifying competencies to enable the successful transition to an AI confident and knowledgeable society. As mentioned above, both Government and a range of stakeholders have an important role to play in setting the parameters of AI operation and leading on the dialogue that is necessary to enable wider societal acceptance.

Q9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

47. The lack of transparency is only appropriate where sensitive data /information is at risk of being exposed. At all other times, transparency should be advocated. If not, this could constitute a breach of ethics and create much discord and distrust in society.

48. AI should follow the example of existing technology sets out there such as cryptocurrency, big and cloud data, banking and medical data. AI is just a connecting tool allowing for the layered learning at rapid cycle rates. If AI is placed into existing technology sets that have already been bedded in, then they should follow whatever acceptable transparency levels already set.

49. To protect intellectual property a degree of black boxing would make sense however anything that would breach a person's human rights should not be allowed to exist inside such a space.

Learning from others

Q11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

50. BIFM does not have any comments as to what particular policy approaches from other countries or organisations would be helpful. In general, as with the world of computing before there are always lessons to be taken away from others' approach to a similar topic. Given the learning curve AI is still going through, there will inevitably be a degree of trial and error. By sharing the learning with others, the gaps will close a lot quicker.

51. looking at other countries and organisations would allow us to benchmark ourselves as a country and to compare our ability to manage the technology and our policies and regulatory framework.

52. Further to this such comparison would allow us to measure how much or little we allow AI to be integrated socially, culturally, economically and the political implications it could have on our society.

Conclusion

53. While the FM industry has grown exponentially over the last decades, it now faces a skills gap due to the UK's demographic change and a lack of applicants with the required aptitude. The uncertainty over a new migration policy adds to that challenge.

54. The skills gap needs a multi-faceted approach to ensure that the UK's FM industry can deliver its potential of a £20 billion uplift to the UK economy by enabling an effective workplace to the UK's businesses as well as retain its status

as one of the most mature and developed FM markets in the world. A flexible migration policy and upskilling people are part of the solution. However, those two approaches will not be sufficient to fill the skills gap and increase productivity.

55. Utilising advances in technology such as AI to enable smart buildings to maximise efficiencies and increase productivity in the workplace are but some of the opportunities that advanced technologies bring. AI will however also bring some further challenges. While upskilling people is part of the solution to the skills gap, lifelong learning will not just educate people about the potential of AI, it will also play an important role in mitigating the challenges of automation and AI and their impact upon our society, by retraining/upskilling people and thus enabling them to retain marketable and much needed skills.

56. The Government should consider the above multi-faceted approach as part of its forthcoming Industrial Strategy when looking at how to upskill people, increase productivity and furthering the potential of AI.

57. In general, Government has a pivotal role to play in facilitating the debate around AI and educating people to enable societal acceptance. In addition, Government can also provide a framework wherein AI can be developed and implemented, governed by principles. Furthermore, Government could lead by example and demonstrate how it is applying AI in its operational functioning.

58. When looking at AI and its role and impact in the FM industry, Government should include commitments in the Industrial Strategy to incentivising the research and use of technologies such as AI and automation. In addition, it could also further encourage the use of building information modelling (BIM) beyond the public sector into the private sector, including involving the private sector in public infrastructure projects. Furthermore, it could encourage more data driven decisions in Government's use of buildings.

8 September 2017

British Standards Institution – Written evidence (AIC0165)

6th September 2017

**Written evidence submitted by the British Standards Institution to the
House of Lords Select Committee on Artificial Intelligence.**

Submission by BSI

1. BSI (the British Standards Institution) is making this submission as the National Standards Body for the United Kingdom. BSI has a public function in support of the UK economy. We bring together stakeholders (including Government, industry and consumers) and facilitate the development of “what good looks like”.
2. BSI would like to make the Committee aware of the opportunities for Government to use standards to deliver its policy objectives in the area of artificial intelligence. We will respond only to the questions of relevance to BSI, as the UK’s National Standards Body.
3. Voluntary standards offer a flexible, adaptive and collaborative alternative to regulation by providing common languages, terminologies, guidelines and good practice developed by and for stakeholders. As the UK’s National Standards Body, BSI operates in accordance with an MOU with the UK Government. Our robust standards development process requires open and full consultation with stakeholders to build consensus based outcomes. This gives standards the legitimacy and degree of market acceptance to be used for public policy purposes.
4. Information Technology standardisation represents a major area of activity for BSI. Our work includes cyber security, Internet of Things, blockchain / DLT (Distributed Ledger Technology), artificial intelligence and virtual/augmented reality.
5. BSI considers artificial intelligence to be an area that could benefit considerably from standardisation. A workshop held by BSI to discuss the subject with stakeholder from the healthcare sector generated evidence that standards will help tackle issues around artificial intelligence. BSI is currently planning a similar workshop for the broader perspective on artificial intelligence later this year.

Question 1: The pace of technological change

6. There is a widespread recognition of the increasing urgency for standardisation in continuously expanding technological systems, which are becoming more complex and interdisciplinary.²¹⁰ Artificial intelligence is potentially one such example. Standards is one of the factors that can accelerate the development of artificial intelligence while helping build trust in the technology and promoting public acceptance.
7. International standards help to deliver global leadership for the UK by accelerating innovation and commercialisation of technologies in areas where the UK has, or is looking to develop, strong R&D capability, such as artificial intelligence.²¹¹ This is in line with the Government Office for Science's recommendation to use standards as market framing policy levers, exploiting 'insights from "living labs" to develop UK standards – setting the global agenda by "showing, not telling"'.²¹²

Question 3: Impact on society

8. As is widely documented, artificial intelligence is likely to bring about profound changes to the way that we live and work. It raises challenging issues, including ethics, process transparency, cyber security, data management and privacy. For all of these topics standards provide a way for the various stakeholders to develop good practice accepted by all parties and to develop trust with consumers.

Question 7: Industry

9. Standards can be of particular value in facilitating innovation by reducing the time to market for new products, promoting the diffusion of innovative products, levelling the innovation playing field between large and small

²¹⁰ See, for example, Tassej, G., (2015). The economic nature of knowledge embodied in standards for technology-based industries. In C. Antonelli & A. N. Link, eds. Routledge Handbook of the Economics of Knowledge. New York: Routledge, pp. 189–208.

²¹¹ There is an increasing body of evidence of the role of standards in supporting and accelerating innovation. See, for example, CEBR (2015). The economic contribution of standards to the UK economy; Blind, K. (2013). The impact of standardization and standards on innovation. Nesta Working Paper No. 13/15. TU Berlin, Rotterdam School of Management and Fraunhofer FOKUS; and Blind, K., Jungmittag, A. and Mangelsdorf, A. (2011). The Economic Benefits of Standardization. DIN Germany Institute for Standardization. Berlin, Germany.

²¹² Government Office for Science (2017). Technology and Innovation Futures, London.

companies, and facilitating the inter-operability in network industries thus creating the environment for the development of new products.

Questions 8 and 9: Ethics

10. BSI would like to bring to the Committee's attention the existing guidance that was facilitated by BSI. BS 8611:2016 "Guide to the ethical design and application of robots and robotic systems" was written by a BSI committee of experts that included scientists, academics, ethicists, philosophers and users. It provides guidance on potential hazards and protective measures, including ethical hazards arising from the growing number of robots and autonomous systems being used in everyday life.
11. The standard also provides guidelines to eliminate (or reduce to an acceptable level) the risks associated with these ethical hazards.
12. The guide addresses many of the questions raised in the call for evidence and BSI would be pleased to discuss its content with the Committee.
13. The subject of transparency provoked much debate in the recent workshop facilitated by BSI for the healthcare sector. Transparency is critical for highly regulated areas such as healthcare. However, stakeholders also highlighted that, due to legislative requirements (e.g. data protection) and the need to protect intellectual property, complete transparency may not be possible.
14. Standards have the potential to develop best practice on transparency that can meet the needs of developers and users of artificial intelligence, whilst meeting regulatory/legislative requirements.

Question 10: The role of Government

15. Voluntary standards offer a flexible, adaptive and collaborative alternative to regulation by providing common languages, terminologies, guidelines and good practice developed by and for stakeholders. As the UK's National Standards Body, BSI operates in accordance with an MOU with the UK Government. Our robust standards development process requires open and full consultation with stakeholders (including Government, industry and consumers) to build consensus based outcomes. This gives standards the legitimacy and degree of market acceptance to be used for public policy purposes.

16. Standards can also be used by Government to support regulation by documenting a method for businesses to comply with regulation.
17. Problems can arise when regulatory systems fail to deal with new technologies or to adapt quickly. This is highlighted in research conducted by the Innogen Institute at the University of Edinburgh for BSI. *“Proportionate and adaptive governance of innovative technologies”* recommends the adoption of guidelines and standards, alongside and/or in place of regulation, to promote the safe adoption of emerging technologies whilst maintaining a focus on innovation.²¹³

Background on BSI

BSI is the UK’s National Standards Body, incorporated by Royal Charter and responsible independently for preparing British Standards and related publications and for coordinating the input of UK experts to European and international standards committees. BSI has over 115 years of experience in serving the interest of a wide range of stakeholders including government, business and society.

BSI represents the UK view on standards in Europe (via the European Standards Organizations CEN and CENELEC) and internationally (via ISO and IEC). BSI has a globally recognized reputation for independence, integrity and innovation ensuring standards are useful, relevant and authoritative.

BSI is responsible for maintaining the integrity of the national standards-making system not only for the benefit of UK industry and society but also to ensure that standards developed by UK experts meet international expectations of open consultation, stakeholder involvement and market relevance.

British Standards and UK implementations of CEN/CENELEC or ISO/IEC standards are all documents defining best practice, established by consensus. Each standard is kept current through a process of maintenance and review whereby it is updated, revised or withdrawn as necessary.

Standards are designed to set out clear and unambiguous provisions and objectives. Although standards are voluntary and separate from legal and regulatory systems, they can be used to support or complement legislation.

²¹³ Tait, J., Banda G. (2016). Proportionate and adaptive governance of innovative technologies. The role of regulations, guidelines and standards; and Tait, J., Banda G., Watkins, A. (2017). Proportionate and adaptive governance of innovative technologies (PAGIT) – Phase 2, available from: <https://www.bsigroup.com/research-pagit-uk/>

Standards are developed when there is a defined market need through consultation with stakeholders and a rigorous development process. National committee members represent their communities in order to develop standards and related documents. They include representatives from a range of bodies, including government, business, consumers, academic institutions, social interests, regulators and trade unions.

Further Information

Steve Brunige
Head of Government & Industry Engagement

Tim McGarr
Market Development Manager (Digital)

6 September 2017

British Standards Institution – Supplementary written evidence (AIC0231)

BSI (the British Standards Institution) welcomes the Committee's evidence gathering on AI and would like to provide an update on developments in standardization since our response to the call for evidence.

In the recent report "Growing the AI Industry in the UK²¹⁴" by Professor Dame Wendy Hall and Jérôme Pesenti, the UK is viewed as a centre of expertise for AI. The report notes a need for standards that provide guidance on how to explain decision-making and processes enabled by AI.

BSI is bringing together UK stakeholders to build market leading best practice around AI. On the 20th October we facilitated a joint roundtable discussion with Nesta and techUK to determine the need for cross-sector standards in AI. One of the key concerns raised was the abundance of sub-standard initiatives on the use of AI that are damaging consumer and business confidence in this emerging technology.

We are now working with Nesta to raise awareness of these issues and help industry and consumers identify the best quality AI products and services during the procurement process. A white paper will be produced from the findings of this roundtable and this will complement BSI's strategy for the development of AI standards.

At an international level, many countries - most notably China and US - see standards as a tool for delivering industry leadership and are already developing standards in AI. For example, in the US, the IEEE (Institution of Electrical and Electronics Engineers) has been exploring the ethical side of AI since 2016. BSI has held preliminary discussions with IEEE on AI and will continue to explore potential collaborations on AI standardization.

An international standards committee has now been formed and is already moving towards commencing foundational standards. BSI plans to take an active role in the work of this committee to deliver the standards necessary to support the development of the UK's AI industry.

BSI would be pleased to provide further evidence to the Select Committee. If you require any further information, please do not hesitate to contact us.

Steve Brunige

Tim McGarr
Market Development Manager (Digital)

²¹⁴ <https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk>

Head of Government & Industry
Engagement

15 November 2017

BSA The Software Alliance – Written evidence (AIC0153)

1. BSA | The Software Alliance (“BSA”) welcomes this opportunity to respond to the UK House of Lords Select Committee on Artificial Intelligence’s (“Select Committee”) call for evidence on artificial intelligence (“AI”). BSA members include many of the world’s leading suppliers of software, hardware, and online services to organisations of all sizes and across all industries and sectors.²¹⁵ BSA members have made significant investments in developing innovative AI solutions for use across a range of applications. As leaders in AI development, BSA members have unique insights into both the tremendous potential that AI holds to address a variety of social challenges, and the types of governmental policies that can best support AI innovation and its responsible use.
2. The range of potential benefits from the smart use of AI is vast. As we describe further in the next section of this paper, AI solutions are already leading to improvements in healthcare, advances in education, more robust accessibility tools, stronger cybersecurity, and increased business productivity and competitiveness. AI also has the potential to generate substantial economic growth and enable governments to provide better and more responsive government services while addressing some of their most pressing societal challenges, as we discuss in Parts II and III.
3. Given the UK’s role as a global leader in technology innovation and development, it is well-positioned to capitalize on these benefits. To maximize the potential of AI, however, the UK Government must adopt a constructive policy framework to support a positive trajectory for AI, which we describe in Part IV of these comments. In pursuing such policies, policymakers should pursue a fact-based, pragmatic approach to regulation, avoiding a one-size-fits-all solution. We see the Select Committee’s call for evidence in this case as an important and valuable contribution towards this goal.

I. AI Is a Tool to Improve Decision-Making, Not a Substitute

4. The Select Committee asks respondents to state how they define AI. We welcome the Select Committee’s question, since there is significant confusion in the popular media about what AI really is and what it does. In brief, virtually all AI systems at their core assist in the analysis of data to find connections that improve the quality and accuracy of human decision-making. The data analytics driving AI systems use sophisticated

²¹⁵ BSA’s members include: Adobe, ANSYS, Apple, Autodesk, Bentley Systems, CA Technologies, CNC/Mastercam, DataStax, DocuSign, IBM, Intel, Intuit, Microsoft, Oracle, salesforce.com, SAS Institute, Siemens PLM Software, Splunk, Symantec, The MathWorks, Trend Micro, Trimble Solutions Corporation, and Workday.

algorithms implemented through software tools. An algorithm, in turn, is a set of instructions that collects inputs and provides an output in a systematized method. The algorithms used in AI are often particularly well-suited to analyzing massive volumes of data from many different sources, and reflecting variables that may interact in complex and unexpected ways. AI algorithms enable technological solutions that enhance perception, learning, reasoning, and decision-making aimed at improving the ability of people to solve complex and challenging problems.

5. Although some AI systems are sometimes described as “autonomous,” the fact is that very few AI systems operate independently of human direction, and most AI systems aid rather than replace human decision-making. It is also important to keep in mind that AI systems can be used in an almost unimaginably wide variety of different contexts, and to improve an exceedingly diverse array of business and consumer experiences across a range of applications and devices. As a result, one set of broad rules seeking to regulate all forms of AI will almost always be over-prescriptive, chilling or even prohibiting beneficial uses of AI in some areas while possibly failing to adequately regulate in others.

II. The Benefits of AI for Individuals and Enterprises

6. AI systems already provide enormous benefits to people and enterprises, including public-sector entities, and they have the potential to generate even greater benefits in the years ahead. Although the range of AI applications is too vast, and evolving too quickly, to summarize here, we highlight a few areas in which AI solutions are already transforming important sectors of the economy and society and, in doing so, are providing concrete, tangible benefits for both people and enterprises.

Benefits for Individuals

AI provides concrete, tangible benefits for individuals across a range of contexts—including healthcare, education, and accessibility.

- **7. In healthcare**, AI technologies are already providing solutions that are helping save lives. A 2016 Frost & Sullivan report predicts that AI has the potential to improve health outcomes by 30 to 40 percent.²¹⁶ AI is helping fuel these improved health outcomes not by replacing the decision-making of healthcare professionals, but by giving these professionals new insights and new ways of analysing and understanding the health data to which they have

²¹⁶ See *From \$600 M to \$6 Billion, Artificial Intelligence Systems Poised for Dramatic Market Expansion in Healthcare*, Frost & Sullivan (Jan. 5, 2016), at <https://ww2.frost.com/news/press-releases/600-m-6-billion-artificial-intelligence-systems-poised-dramatic-market-expansion-healthcare>.

access. For example, AI tools are powering machine-assisted diagnosis and surgical applications used to improve treatment options. Image recognition algorithms are helping pathologists more effectively interpret patient data, thereby helping physicians form a better picture of patients' prognosis. These improvements are helping improve diagnostic and surgery outcomes, saving countless lives. This ability of AI to process and find patterns in vast amounts of data from disparate sources is also driving important progress in biomedical and epidemiological research.

- **8. In education,** AI tools are changing how schools are run and how educators teach students, including by helping them quickly identify students that need particular attention and giving them the support they need. AI can automate basic activities to assist teachers, such as grading, which allows teachers to further interact and assist students. AI software adapts to student needs, enabling personalized learning, and identifies areas where courses need to improve.
- **9. In the accessibility context,** AI solutions are powering devices and software programs that improve and enrich the lives of people with disabilities. For instance, AI tools are helping people with vision-related impairments to interpret and understand photos, other visual content, and websites on the Internet, and even to navigate their physical surroundings. Microsoft recently released an intelligent camera app, for instance, that uses a smartphone's built-in camera functionality to describe to low-vision individuals the objects that are in front of them.²¹⁷

Benefits for Enterprises

10. AI also provides opportunities to increase business competitiveness and innovation. For instance, AI can accelerate production capabilities through more reliable demand forecasting and increased flexibility in operations and supply chains.²¹⁸ It can create smarter, faster, cheaper, and more environmentally-friendly production processes that increase worker productivity, improve product quality, lower costs, and improve worker health and safety.²¹⁹
11. In all of these contexts, AI is useful not because it replaces humans, but rather because it enables humans to focus on tasks to which they add the

²¹⁷ See Microsoft, *Seeing AI*, at <https://www.microsoft.com/en-us/seeing-ai/>.

²¹⁸ U.S. National Science and Technology Council and Networking and Information Technology Research and Development Subcommittee, *National Artificial Intelligence Research and Development Strategic Plan*, 8 (Oct. 2016), at https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf.

²¹⁹ See *id.*

greatest value. The cybersecurity context offers a prime example of this benefit. AI tools are able to monitor networks and identify aberrations that warrant further attention by network administrators, helping these security professionals hone in on issues that represent the most significant cyber threats. AI programs can also automatically isolate suspicious network traffic until security professionals can examine it, preventing the spread of malware in a network even if the malware is successful in breaking through network defenses.²²⁰ These learnings are then used to further improve future detection as well as to improve existing software.

12. Governments, too, can leverage AI tools to become smarter, more efficient, less costly, and more responsive in how they provide public services. AI tools can enable public agencies to automate routine tasks, reinvigorate citizen engagement through new communication portals, and securely synthesize vital health, economic, and public data.²²¹ For example, in the United States, the Cincinnati Fire Department has optimized its medical emergency response process using an AI-based system, enabling the Department—which responds to around 80,000 medical emergencies each year—to more strategically position its

²²⁰ For example, IBM's Watson for Cyber Security is a cybersecurity tool that can analyze 15,000 security documents per day—a rate essentially impossible for any individual to achieve. Watson's data processing capabilities enable analysts to more quickly identify incidents that require human attention. See IBM, *IBM Delivers Watson for Cyber Security to Power Cognitive Security Operations Centers*, (Feb. 13, 2017), at <https://www-03.ibm.com/press/us/en/pressrelease/51577.wss>; Jason Corbin, *Bringing the Power of Watson and Cognitive Computing to the Security Operations Center*, Security Intelligence (Feb. 13, 2017), at https://securityintelligence.com/bringing-the-power-of-watson-and-cognitive-into-the-security-operations-center/?cm_mc_uid=70595459933115020631816&cm_mc_sid_50200000=1503364089&cm_mc_sid_52640000=1503365578. Splunk uses a similar model, with machine-learning algorithms conducting real-time analysis and processing of massive volumes of data from all sensors on a network to identify anomalies, feeding visualization tools that help network administrators efficiently triage security incidents. See Braue, David, "Machine learning key to building a proactive security response: Splunk," *CSO Online*, (August 20, 2015), <https://www.cso.com.au/article/582483/machine-learning-key-building-proactive-security-response-splunk/>. Microsoft's Windows 10 Anniversary Edition introduced AI-driven capabilities for automatically isolating suspicious network traffic pending adjudication by network administrators. See Hallum, Chris, "Defense Windows clients from modern threats and attacks with Windows 10," Channel 9 video content (October 6, 2016), <https://channel9.msdn.com/events/Ignite/2016/BRK2135-TS>; "Intelligent Security: Using Machine Learning to Help Detect Advanced Cyber Attacks," <https://www.microsoft.com/en-us/security/intelligence>.

²²¹ See Kriti Sharma, *Artificial intelligence can make America's public sector great again*, Recode (Jul. 14, 2017), at <https://www.recode.net/2017/7/14/15968746/artificial-intelligence-ai-federal-government-public-sector>.

personnel and reduce both response times and the overall number of runs.²²²

III. The Economic and Social Impacts of AI

13. Although most of us will experience the benefits of AI in the coming years at an individual level, AI is also set to have significant macroeconomic and societal benefits as well. Experts predict that applications of AI technologies could grow the global economy by between \$7.1 to \$13.17 trillion over the next eight years.²²³ And as the UK Government has recognized, the application of “big data,” which fuels many AI technologies, is estimated to generate £240 billion in cumulative benefits to the UK between 2015 and 2020.²²⁴ Similarly, in the United States, the market for AI technologies that analyze unstructured data is expected to reach \$40 billion by 2020, creating over \$60 billion worth of productivity improvements each year.²²⁵

14. The broader social impacts of AI can take many forms. For instance, AI can improve individual and public safety by helping people in anticipating and responding to dangerous situations. AI-powered distributed sensor systems and pattern understanding of environment conditions can detect when the probability of major infrastructure disruptions increases significantly and help enterprises adapt operations as needed to respond to disruptions even before they occur.²²⁶ AI can also improve

²²² See Kevin C. Desouza, Rashmi Krishnamurthy, and Gregory S. Dawson, *Learning from public sector experimentation with artificial intelligence*, Brookings Institution (June 23, 2017), at <https://www.brookings.edu/blog/techtank/2017/06/23/learning-from-public-sector-experimentation-with-artificial-intelligence/>.

²²³ See *Artificial Intelligence in Canada: Where Do We Stand?*, Information and Communications Technology Council 2 (Apr. 2015), at <https://www.ictc-ctic.ca/wp-content/uploads/2015/06/AI-White-paper-final-English1.pdf> (citing *Disruptive technologies: Advances that will transform life, business, and the global economy*, McKinsey Global Institute (May 2013), at <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/disruptive-technologies>).

²²⁴ See UK Government, Office for Science, *Artificial intelligence: opportunities and implications for the future of decision making 9* (2015) [hereinafter *UK Government Office for Science Report*], at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf.

²²⁵ Software more generally delivers a total value-added (direct, indirect, and induced) GDP of €910 billion — 7.4 percent of the EU28 total. In the UK alone, software industry in the UK directly contributed €65.3 billion to the economy — the highest of any country in the EU and almost 3 percent of UK GDP. See BSA, *Software: A €910 Billion Catalyst for the EU Economy* (2016), at <http://softwareimpact.bsa.org/eu/>.

²²⁶ See *National Artificial Intelligence Research and Development Strategic Plan*, *supra* note 4, at 19 (citing Bela Genge, Christos Siaterlis, and Georgios Karopoulos, *Data fusion-based anomaly*

transportation and mobility. AI already powers tools that help people gain new insights into traffic flows, travel times, and optimal routes, increasing energy efficiency and reducing pollution levels and commute times. For example, large cities have begun to leverage the type of responsive dispatching and routing technologies used by popular ride-sharing services by linking them with scheduling and tracking software for public transportation to provide just-in-time access to public transportation that can often be faster, cheaper and, in many cases, more accessible to the public.²²⁷ Using sensors and cameras in the road network, they can also optimize traffic light timing to improve traffic flow and to help with automated enforcement.²²⁸

15. BSA recognizes that AI also raises concerns relating to its potential impact on employment. Although it is important that both industry and governments address these concerns, BSA agrees with the UK Government Office for Science that, despite the potential for AI to displace certain jobs, “we should expect that new types of job[s] will emerge” as a result of AI and that “new industries may emerge and grow as productivity gains lead to higher incomes and declining costs.”²²⁹
16. BSA members have already begun helping workers transition to jobs that require different skills that will complement AI systems. BSA members offer a number of high-tech and business training programs, including at high school level. There are also a number of programs that explore new innovative approaches to using AI to connect skills to available

detection in networked critical infrastructures, 43rd Annual IEEE/IFIP Conference on Dependable Systems and Networks Workshop (DSN-W), 2013).

²²⁷ Executive Office of the President National Science and Technology Council Committee on Technology, *Preparing for the Future of Artificial Intelligence* 23 (Oct. 2016) (citing Stephen F. Smith, *Smart Infrastructure for Urban Mobility*, presentation at AI for Social Good workshop, Washington, DC (June 7, 2016)) at https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/p_reparing_for_the_future_of_ai.pdf.

²²⁸ Peter Stone, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyan Krishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller, *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel 22*, Stanford University (Sept. 2016) (internal citations omitted).

²²⁹ See UK Government Office for Science Report, *supra* note 10, at 12.

opportunities, moving the dialogue from job experience to skills that are needed.²³⁰ These initiatives illustrate just some of the ways in which AI-based employment concerns can be meaningfully addressed.

IV. How the UK Government Can Support a Positive Trajectory for AI

17. As a global leader in science, technology, and research, and benefitting from a skilled, well-educated labour force, the UK is advantageously positioned to seize the opportunities of AI. The UK is fortunate to have a vibrant culture of entrepreneurialism, ranking among the top five most innovative countries in the world, according to the Global Innovation Index.²³¹ The UK also has fostered much of the scientific research underpinning AI technology at its world-class academic institutions.²³² In short, the UK starts from a position of strength in terms of capturing the future benefits of AI.
18. To maintain this edge, however, and to ensure that British citizens reap the full potential benefits of AI, the UK would do well to adopt a

²³⁰ See, e.g., Blue, A., *How LinkedIn is Helping Create Economic Opportunity in Colorado and Phoenix*, <https://blog.linkedin.com/2016/03/17/how-linkedin-is-helping-create-economic-opportunity-in-colorado-and-phoenix> (Mar. 17, 2016); *Why Microsoft and the Markle Foundation are Working Together to Connect Workers with New Opportunities in the Digital Economy*, <https://www.markle.org/microsoft>. IBM, for instance, has established Pathways in Technology Early College High Schools (P-TECH Schools). P-TECH schools are innovative public schools that offer students the opportunity to earn a no-cost associates degree within six years in fields such as applied science and engineering—and to acquire the skills and knowledge necessary to pursue further educational opportunities or to step easily into well paying, high-potential informational technology jobs. IBM designed the P-TECH model to be both widely replicable and sustainable as part of an effort to reform career and technical education. See IBM, *IBM and P-TECH*, at <https://www-03.ibm.com/press/us/en/presskit/42300.wss>. Likewise, Salesforce has been offering free high-tech and business skills training through Trailhead, an online program that provides students business and technology training, with the goal of preparing them for the nearly 140,000 Salesforce-based jobs that the company expects will be created between 2015 and 2020. See Gavin Mee, *Guest Blog: Gavin Mee, Salesforce - Evolving tech means change in digital skills*, TechUK (Apr. 26, 2017), at <https://www.techuk.org/insights/opinions/item/10695-guest-blog-gavin-mee-salesforce-evolving-tech-means-change-in-digital-skills>.

²³¹ *The Global Innovation Index 2017: Innovation Feeding the World 20* (2017), at <https://www.globalinnovationindex.org/>.

²³² See UK Government Office for Science Report, *supra* note 10, at 8.

policy framework that promotes AI innovation and does not impose barriers that stifle AI uptake. Key elements of such an approach include the following:

19. **Support broad deployment and continued innovation of AI through investments and leading by example.** Maximizing the potential of AI will require substantial investments by both the public and private sectors. The UK should invest directly in AI research, development, and deployment, and further support its growth by providing incentives for private-sector investments in AI—for instance, through strategic tax credits for private-sector R&D. The UK Government could also help demonstrate AI’s potential benefits by investing in innovative AI implementations in the public sector. Broader deployment of AI will also lead to further innovation and advancements across multiple sectors.
20. **Support voluntary, industry-based efforts to promote accountability.** As the Select Committee’s questions recognize, certain uses of AI can raise ethical, fairness, due process, or related concerns. There have been several recent private-sector efforts to address these concerns by promoting accountability of AI through approaches that would provide a broader understanding of how certain AI systems operate but do not otherwise require disclosure of confidential business or other proprietary information. The UK government should support such efforts, which are far more likely to address relevant concerns than broad, one-size-fits-all disclosure mandates that may pose privacy and other concerns, while not addressing the primary question of increasing public understanding of these systems.
21. **Avoid barriers to cross-border data transfers.** AI systems use computational analysis of data to uncover patterns and draw inferences. This data may originate from many sources located in multiple jurisdictions, making it imperative that data can move freely across borders. Rules that limit cross-border data transfers invariably limit the insights and other benefits that AI systems can provide. The UK should support strong trade commitments to facilitate data flows, including commitments against mandates to locate computer facilities domestically.
22. **Support exceptions to copyright liability for text and data mining.** Because AI systems analyze data exclusively to draw inferences—and do not seek to publish or otherwise exploit the expressive content that may inhere in such data—national copyright systems should include a broad exception to copyright liability for text and data mining by any user with

lawful access to content. The UK's support for such an exception could help reaffirm the country's leadership position on AI policy.

23. **Ensure that any regulation is technology neutral.** Given the tremendous promise that AI technologies offer, it is important to avoid stunting their growth through overly broad *ex ante* regulation. The UK's overall approach to regulation should be sectoral and technology neutral. Thus, policymakers should refrain from imposing broad-based regulatory constraints on this still-developing technology. The UK Government should only consider carefully and narrowly tailored policy responses targeting actual issues of concern in the exceptional situations in which such responses are needed, rather than adopting regulations targeting the underlying technology. This approach will be most likely to yield optimal outcomes for the public.
24. **Prepare the UK workforce for the jobs of the future.** According to a recent report by the World Economic Forum, 65% of today's children will hold jobs that have not been invented yet.²³³ It is very important that the UK Government work with the private sector to develop a national strategy for ensuring that UK workers continue to have the skills necessary to thrive in the new data economy. We urge the UK Government to work with the private sector to develop a robust strategy to boost efforts to foster education at all levels in the STEM fields, such as computer and data science and software engineering, as well as to improve core competencies taught to grade and high school children, including problem-solving skills. Strategies should also focus on technical training provided outside traditional college programs.

V. **Conclusion**

The UK Government is poised to help usher in a new wave of technology that offers direct, groundbreaking benefits, with transformative economic impact for its citizens, its businesses, and itself. Realizing these benefits will, however, require a thoughtful and measured regulatory framework that accounts for the technological reality and variety of AI technologies, supports robust and free data flows, and facilitates meaningful investments in further AI research and development.

6 September 2017

²³³ Vincenzo Spiezia, *This is what coachmen from the 1920s can tell us about robots and jobs*, World Economic Forum (Jul. 18, 2016), at https://www.weforum.org/agenda/2016/07/this-is-what-coachmen-from-the-1920s-can-tell-us-about-robots-and-jobs/?utm_content=bufferb59f7&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer.

Dr Aysegul Bugra, Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis and Dr Valentina Rita Scotti – Written evidence (AIC0051)

Dr Aysegul Bugra, Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis and Dr Valentina Rita Scotti – Written evidence (AIC0051)

Response to the Call for Evidence of the House of Lords Select Committee on Artificial Intelligence

The Respondents (in alphabetical order):

Dr. Aysegul Bugra, Director of NASAMER (Dr. Nusret-Semahat Arsel International Business Law Implementation and Research Center); Assistant Professor, Koç University

Matthew Channon, Lecturer and Ph.D Candidate, University of Exeter

Dr. Ozlem Gurses, Associate Professor, King's College

Dr. Antonios Kouroutakis, Assistant Professor, Instituto de Empresa

Dr. Valentina Rita Scotti, Post-doctoral Research Fellow, Koç University

The respondents submit this written evidence in their personal capacity.

Yours sincerely,

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1. AI is currently rather at an experimental stage, but at the same time its progress is rapid. Its emergence in our daily life will be incremental and will strengthen its presence progressively. Thus, the mark of AI will be evaluated ex post (like the birth of the modern state with the Westphalian Treaties). Without doubt, the strict regulation of AI will hinder the development of AI, but the competition between the divergent regulatory regimes (ie in the EU, US, and in China) will accelerate its development. We foresee a development of AI in several fields. In the military sector, killer-robots and robot-interrogators will begin operating. Unmanned Aircraft Systems (UAS) or drones are already in use, and there exists a growing demand for non-military use of robots in the civil environment for a number of governmental functions like policing, border control, search and rescue, fire-fighting, ground traffic surveillance, and pollution control.

1.2. Furthermore, there is significant development of Advanced Drivers' Assistance systems (ADAs), along with Telematics, which are undoubtedly leading to safer roads and greater understanding of driving patterns, making it

both easier and cheaper to insure vehicles. This alongside the development of Connected and Autonomous Vehicles (CAVs), will have a significant impact on how vehicles are used and owned. This is a trillion dollar market being driven by the government, manufacturers and a number of technological startup companies, with the aim of making roads safer, travel easier whilst saving substantial amounts of money. CAVs are at a relatively early stage, with vehicles currently being tested at Level 3 (SAE International Standard J3016), with projects underway to introduce CAVs to the public (see in particular Volvo's 'Drive me' Project). In the next 20 years it is likely that a number of 'smart cities' will be developed, with CAVs being used as taxis rather than the public owning a car. Car Pooling is already beginning to take up, with a number of companies such as Bla Bla Car (from Axa) entering the market.

1.3. Robots are also used in medical procedures and operations such as in diagnostic systems, robot-assisted surgery and therapy, and rehabilitation systems; industrial robots are used in manufacturing sectors; and finally, there are service robots for personal and domestic use, such as robots that will help the elderly or take over household chores. The new developments will soon include robots for child-caring and sex-bots.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

3.1. One of the most problematic impact is on human employment. Day by day we hear about AI being used for employers to perform the tasks that used to be performed by human employees. This seems to be the greatest concern in the society for obvious reasons. Employers are likely to prefer to employ robots rather than humans as robots are employed more cheaply compared to humans. Moreover, robots are likely to be able to work longer hours than humans and thus, likely to be more productive. The question is whether AI will outperform humans at almost every cognitive tasks? The areas that AI are and will be used is the first issue that should be addressed. Is there going to be a restriction in terms of developing AI in a way replacing human power in employment? The answer to this questions may seem to be straightforward, however, restricting research will not be favoured by researchers. On the other hand, protecting human rights will have to be considered in terms of the employability of humans. Thus, some restrictions in terms of the areas that AI may be employed and the number of AI that will be permitted per employer will likely to be considered.

3.2. One of the most important ways in which the public can be prepared is in education, there are undoubtedly significant issues relating to the public's perception of AI. There is a wariness that AI will cause a threat to the general

public and it is important early on, to ensure that the public are aware of how AI is developed and how it will benefit daily life. Without adequate education, it is unlikely that AI will become popular amongst most of the public.

3.3. The use of AI may also affect the protection of the right to privacy and security, making the introduction of new laws necessary also in these fields for controlling AI activities related to sensitive data and for sanctioning any violations that may occur. Finally, it cannot be ignored that the development of AI and of robotics may produce also the need to legislate about whether they should have legal personality, as was suggested in the European Parliament Committee on Legal Affairs Draft Report (with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) dated 31 May 2016 (hereinafter referred to as the 'EP Report').

3.4. A regulatory intervention will therefore be unavoidable in order to safeguard social cohesion put in danger in case AI will replace human activities without rules aimed at controlling it. This risk will particularly arise with respect to robots that will operate in households and will have direct interaction with people. For instance the emergence of sex-bots designed to replace sexual activities among people may have the following double effect: as they will be provided to their owners for sex without the need for consensus, this may have an impact on the owners' perception of consensus for sex, which consequently may alter their interaction with their potential partners, and in longer term may accordingly also affect the reproductive cycle, due to the fact that people may start finding it easier to buy sex-bots than to search for a partner. Similarly, we submit that specific regulation will also be required regarding household robots that are used in child or elderly-care as they are expected to affect the social skills and attachment patterns of the part of the population that will receive such care.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

4.1. The answer to the first question is very broad. The use of AI may be little more than a minor nuisance if your laptop crashes or gets hacked, or it becomes all the more important if an AI system controls cars or airplanes or automated trading systems. It is questionable if an AI can perform a medical surgery or defend a criminal at a court or decide on a case as judges do. Whilst in some areas humans can easily be replaced by AI some others do not seem to be as open to AI. AI can be programmed for some tasks which are performed more or less always in the same way – as routine tasks. For instance, lifting and carrying heavy materials. They can be designed to carry them in some certain orders, to stop when needed and carry on when needed and delivery at the destination. On the other hand, an AI is unlikely to be able to negotiate an agreement or bargain or enter into any type of discussions in which - as the saying goes - one

conversation leads to another. Disparities are unavoidable in terms of the use of AI.

4.2. Entities such as insurance companies and companies able to generate and process big data will particularly gain as they will be able to foresee consumers' reactions and choices. This will strengthen their share in the market economy at the expense of a competitive market economy. Such potential benefits could be mitigated with the strengthening of the regulatory power of the competition authorities.

4.3 There is nothing to say, however, that all of society cannot benefit significantly from the use of AI. Particularly with regards to CAVs. For example, one particular positive in relation to CAVs is in relation to the environment. Current conventional vehicles cause significant environmental pollution, however, with CAVs the environmental damage caused by vehicles will almost certainly significantly be reduced. Moreover, benefits should be felt throughout the economies with CAVs, by reducing accidents by up to 90 percent, there are significant financial savings to be made, for example in hospitals, where there will be less people injured as a result of road accidents.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

5.1. This is inevitable. Considering the areas that AI is used, it is inevitable that the number of humans who either benefit from the service of AI or whose jobs are replaced by AI will elevate day by day. Humans have already been working with AI and as such areas will enlarge the need for public awareness on AI will be required even more heavily. One of the ways of increasing public awareness is the use of media – especially TV and radio shows by which it is possible to reach vast amount of people at the same time. At schools it may be necessary to introduce classes to teach the impact of AI to school pupils. That may also help them understand the professions which will be open to them and will not be threatened by AI or in which they can work with an AI.

5.2. Some care, however, needs to be taken to ensure that the public do not get negative understanding of AI. For example in relation to CAVs there is potential that every accident that occurs will get widespread media coverage, causing negative public perception. Whilst freedom of media is important, it is equally important that media coverage is accurate, particularly when it is human fault rather than the fault of technology.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

6.1. AI will have a spill-over effect in most industries in the long term; however the ones that will immediately be affected are insurance, manufacturing (some manufacturing tasks will be completely replaced by robots, and in other fields it is expected that humans and robots (cobots) will have a co-operative relationship), healthcare, transport (public and otherwise), start up technology companies, telematics and defense.

With respect to insurance of robots, the below points must be considered:

a. Robots as products and challenges in the product liability insurance sphere

6.2. Robots other than the ones having reached a level of autonomy whereby they could be granted electronic personality would fall under the scope of the Directive 85/374/EEC concerning liability for defective products. The Directive is however arguably not sufficient to cover the damages occasioned by the new generation of robots which may not have reached full autonomy yet which would still interact with their environment in a unique and unforeseeable manner as their adaptive and learning abilities increase. Between products of manufacturers in the traditional sense of the word and fully autonomous electronic persons, these semi-autonomous robotics could nevertheless be insured under product liability insurance policies. Moreover, robots can, for instance, cause moral damages arising from their prolonged interaction with people and liability systems together with the scope of product liability policies may accordingly be matters that would require further analysis.

b. Compulsory insurance

Employer's liability insurance

6.3. Robotics are used in work places and they increasingly replace human beings. Whether they are considered as employees akin to their human counterparts and having similar rights and obligations or as products will potentially depend on their level of autonomy. In several jurisdictions including the UK, employers are required to insure employees against personal injury and death that occur in the course of their employment. The question in the next few years will therefore be how employers' liability insurance will apply in the form of insuring robotics as employees. It is likely that employer's liability insurance will not apply to AI. However, some other types of insurance which have already been developed in the insurance market may do. It will be necessary to identify what type of insurance will be required by the employers who employ AI.

Public liability insurance

6.4. It is likely that an AI will perform its tasks which are programmed by a human. If an AI causes damage to third parties the human in question will need public liability insurance. The EP Report proposed a registry for fully autonomous robots, a compulsory insurance scheme whereby producers or users of robots would have to take out insurance cover, the establishment of a compulsory

insurance fund for any third party loss caused by uninsured robots where donations for the development of robotics could also be made. These proposals would be relevant if the UK seek to attempt to regulate public liability insurance for fully autonomous robots.

Motor Insurance

6.5 As with current conventional vehicles, CAVs will need to be insured on roads and all public places when they are being used to ensure that victims of road accidents are adequately compensated. This presents a significant challenge as liability will often shift away from the driver's insurer to the manufacturer's insurer, causing significant future legal dilemmas. The proposed UK system in the Automated and Electric Vehicles Bill of a 'dual insurance' system, whereby both manufacturer's and driver's insurance fall within one policy is a potential solution to this. There are, however, a number of legal issues which arise, involving cross-border travel, liability chains and insurance scope which will need to be thought of in the future.

c. Fully autonomous robots and life insurance

6.6. An issue that can possibly arise in the future is whether it could be possible to provide insurance cover on the 'life' of artificial intelligence machines. This possibility currently transcends the borders of traditional life insurance policies. On the other hand it is anticipated that the recognition of fully autonomous robots as electronic persons could theoretically give rise to the query of whether, for example, people under the care of such AI machines could have insurable interest based on their ties of affection in the 'survival' thereof and therefore could take out life insurance policies on their lives. A tentative answer, if any, could lie in whether there can theoretically exist a degree of autonomy which could accord the abilities of thinking and hermeneutics to AI machines which is a query yet to be shed light upon. If answered in the affirmative, insurance industry would need to offer policies drafted specifically for insuring the lives of AI machines.

7. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

7.1. Data management and its product shall be "patented" for a certain and limited period of time like the drug patents. The patent of data management will be a win-win model for both the companies and the society, as the companies will have the incentive to manage data, and once they become patent free, then the society as a whole will benefit.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

8.1. Undeniably, the use of services provided by robotics will impact on the employability of people. On the one hand, human workforce will be displaced in vast numbers, but also a reduction of human wages will occur. Notably, Bill Gates had proposed introducing “robot tax” to compensate for the loss of human jobs. This has a potential of discussing a new form of positive rights in the form of an entitlement to the profits from the robots’ workforce.

8.2. Besides the possible challenges to security and employment, there is also the risk that the cohabitation with robots, having unlimited capacity of recording and processing data, multiplies the threats to the protection of sensitive information. The possibility that robots may modify, share or hide some personal data has a relevant impact in several fields, ranging from intellectual property (already challenged by the emergence of ITC) to the protection of sensitive States data and of personal data. The latter meaning that a violation of the right to privacy could occur. The robotics designers may have to develop procedures that will require valid consent to be given before the recording of personal data by AI machines.

8.3. Finally, a quite problematic use of AI is in times of war. For instance, killer robots which are in practice lethal autonomous weapons - like weapons of mass destruction - may pose a threat to humanity and violate internationally recognized rules in times of conflict. Please refer to the response provided for Q11 for other countries’ and institutions’ approach to ethical implications of AI.

8.4. There are some perceived ethical issues in relation to CAVs with the ‘Trolley Problem’, the decisions to be made by the vehicle in determining who to drive into if the vehicle cannot stop. This ethical dilemma is largely overplayed, as it is often argued that the problem should not suffice if CAVs are developed safely.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called ‘black boxing’) acceptable? When should it not be permissible?

9.1. Black boxing safeguards tech privacy, however there is a need to strike a balance when a compelling state interest (such as in the case of the occurrence of a criminal offense, terrorism etc) requires transparency. The battle between black boxing and the state interest was revealed during the recent case before the US courts regarding the Syed Farook case, who was a terrorist using an iphone while the Federal Government wanted to access his phone.

It seems that there is a need for a special process before an independent tribunal that shall examine such cases in order to strike the appropriate balance between the protection of tech secrets and the state interest for law enforcement.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

10.1. There are aspects such as the process of data, the black boxing, and insurance of robots that need regulation, as existing laws need to be adjusted and tailored to the challenges arising from AI. All the stakeholders must be represented and engaged in any effort toward regulation of AI, namely NGOs (representing consumers and disabled), municipalities, black box producers, autonomous vehicle manufacturers etc.

10.2. The UK Government's response to CAVs in particular has been positive, particularly with regards to insurance regulation. It is clear that the Government is taking on board views of all to ensure effective insurance regulation and this is seen in the Automated and Electric Vehicles Bill.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

11.1. One of the questions that will have to be tackled in the future is whether autonomous robots shall have 'electronic personality' and should be registered. These issues were addressed in the EP Report. Moreover the [European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics \(2015/2103\(INL\)\)](#) highlighted the points of liability, intellectual property rights and the flow of data, ethical principles, research and innovation, standardisation-safety-security, human repair and enhancement, and environment.

Several State jurisdictions, such as the USA, Japan, China and South Korea are considering, and to a certain extent have already taken regulatory actions with respect to robotics. For instance South Korea released a robot ethics charter in 2007 and approved the Intelligent Robot Development and Supply Promotion Act in 2008. Furthermore the German Federal Ministry of Transport and Digital Infrastructure published a [Report](#) on 20th June 2017 on the ethics of Automated and Connected Cars.

2 September 2017

Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams and Professor Robert Fisher – Written evidence (AIC0029)

Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams and Professor Robert Fisher – Written evidence (AIC0029)

[Submission to be found under Professor Robert Fisher](#)

Dr Mercedes Bunz – Written evidence (AIC0048)

Submission to House of Lords Select Committee on Artificial Intelligence

Dr. Mercedes Bunz, University of Westminster

My background and expertise

My research explores the effects of algorithms on knowledge. I studied recent AI developments for the book 'The Internet of Things' (Polity, Nov 2017, co-authored with Graham Meikle) and analysed this topic in a previous book 'The Silent Revolution: How Digitalization Transforms Knowledge, Work, Journalism and Politics without Making Too Much Noise' (Palgrave Macmillan 2014). As the former technology reporter of *The Guardian*, I also have significant knowledge of start-up culture and have just become member of the Internet of Things Working Group organized by the *British Standards Institute*.

Summary of evidence

AI creates a new data paradigm. To advance this kind of AI, the role of government can make quite a difference. By processing data AI programs *create knowledge* i.e. they take over knowledge tasks and even make decisions. To develop and train this kind of AI, large amounts of data are needed in sufficient size and quality. For the creation of this data, the decisions of government are vital. Policy initiatives should: A) ensure that the UK has a strategy for the creation of big datasets in high quality; B) identify sectors in which the creation of those datasets is especially desirable; C) create incentives for businesses to open and share their datasets; D) foster data sovereignty to minimize UK's algorithmic dependence on U.S. products.

1. Datasets are crucial to train AI. A new programming technique known as '*neural networks*' created the current breakthrough in AI technology. Instead of programming rules that should be applied, a neural network *infers rules of categorization* by analyzing correctly labelled data records in the thousands. To train neural networks in this mode of 'deep learning', computer scientists depend on large datasets. AI's ability to recognize objects in images and videos, for example, has evolved with ever larger image-language datasets on which the objects or actions displayed are correctly named. In the recent past, the UK ran an image recognition challenge based on a dataset known as *PASCAL Voc* (2005 – 2012)²³⁴, which allowed computers to train on about 20,000 annotated images. This was soon being surpassed by datasets created in the U.S. such as the

²³⁴ The PASCAL Visual Object Classes Homepage <http://host.robots.ox.ac.uk/pascal/VOC/> was organised by Mark Everingham (University of Leeds), Luc van Gool (ETHZ, Zurich), Chris Williams (University of Edinburgh), John Winn (Microsoft Research Cambridge) and Andrew Zisserman (University of Oxford).

*Flickr30k*²³⁵ set with over 30,000 pictures focusing on people and everyday activities, and Microsoft's *MS COCO* consisting of 300,000 images offering multiple objects per image. Around 2009, the up-to-date largest dataset was created with the University of Stanford's *ImageNet*, which provided over one million images with annotations. *To train and test AI, datasets must be huge.*

2. AI creates a new data paradigm. As the report '*The big data dilemma*'²³⁶ by the Science and Technology Committee has documented, the UK government is aware of the importance of Open Data and of opening government data. In times of AI, however, this data is not just useful for informing businesses and for allowing start-ups to create new services. Datasets are now also essential to train AI that then can create further knowledge. As AI needs to be trained, only areas for which datasets are available can advance. This is making *a data agenda* necessary. AI has created a new data paradigm – from *public infrastructure* to the *creation of knowledge*.

3. The creation of datasets is expensive. Financially, they depend on research funding or have mostly been created by large corporations such as Google, Facebook etc. Data is available everywhere, but datasets to train AI need to be 'cleaned', i.e. labelled and corrected. In the past, AI research projects have turned to Amazon's micro-task marketplace Mechanical Turk.²³⁷ *ImageNet* employed at its peak 48,940 people in 167 countries that sorted and labelled nearly a billion images downloaded from the internet. According to its former director, Stanford Professor Fei-Fei Li, it was at one point one of Mechanical Turk's biggest employers. Its usage in the *ImageNet Challenge* led to massive breakthroughs in AI-driven image recognition. This is a clear indication that data and the creation of datasets and even more their constant update is as laborious and costly as it is central to developing and training AI. *The creation of datasets, laborious and expensive, is essential to advance AI.*

4. A lot of data is siloed and proprietary. Citizens create data every day in public and commercial environments. However, this data is often out of reach for start-ups and programmers. While programmers do know that a needed dataset exists in another company, it is proprietary and cannot be used to train or update their AI. As the Royal Academy of Engineering stated in its 2015 *Connecting Data* report quoted by the Select Committee: 'Much potentially valuable data remains locked away in corporate silos or within sectors'.²³⁸ New developments such as

²³⁵ Flickr30K is hosted by the University of Illinois <https://illinois.edu/fb/sec/229675>.

²³⁶ Science and Technology Committee, House of Commons, *The big data dilemma: Fourth Report of Session 2015–16*, 10 February 2016.

²³⁷ As described in: Karpathy, A. & Fei-Fei, L. (2015) 'Deep visual-semantic alignments for generating image descriptions', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3128–37.

²³⁸ Science and Technology Committee, House of Commons, *The big data dilemma: Fourth Report of Session 2015–16*, February 2016, p. 25, par 53).

the internet of things will mean that more and more data of citizens will be collected in that proprietary manner. 'Data portability' – allowing individuals to re-use their personal data – and voluntary programs such as UK's 'midata' initiative support this. But individual portability will not be sufficient to collect a dataset that allows the creation of knowledge and businesses. The UK needs a strategy to actively create big data, especially in areas of government interest such as healthcare, transport, science and education. Citizens creating data is of value. The data of 1.6 million patients of the Royal Free Hospital that was given to the company DeepMind has made headlines for breaching UK data law, while the immense value of that dataset has been overlooked. *The creation of open datasets should be actively pursued. Specific areas of interest should be identified in a data agenda.*

5. AI frameworks are often bought as a service from U.S. companies. As UK companies and research institutions experience difficulties to get their hands on big UK data, they are struggling to train their own AI-driven programs. This is especially the case in the field of natural language processing, in which U.S. companies got a head start with the collection of data. Many of UK's digital businesses working with conversational technology such as chatbots bought them from one of the five U.S. companies dominating this market: Google, IBM's Watson, Amazon's Lex, Microsoft, or Facebook.²³⁹ Their chatbot frameworks provide natural language processing abilities that are then further specified by UK companies for specific task by adding on 'domain knowledge'. This can mean that UK services who use U.S. chatbots send their data back to U.S. servers including in sensitive areas such finance (UK banks using chatbots) or healthcare. *Data sovereignty should be a subject addressed in the data governance framework.*

6. AIs are prone to be biased. The Science and Technology Committee has noted this in their '*Robotics and artificial intelligence report*²⁴⁰ as being of ethical and legal concern. In face of the rapid applications of AI in a range of sensitive sectors such as news production (the BBC and the Press Association are currently working on AI projects) or healthcare (Deepmind, Babylon Health and others) a government strategy to address potential bias is needed. AI learns to categorize by looking for patterns, and can easily amplify existing biases resulting from biased training data or by an insufficient training of the AI. *Recommendations for the creation of datasets could help to minimise bias.*

7. Ensuring privacy and consent is another important ethical concern noted in the Science and Technology Committee's report. New approaches such as 'Differential Privacy' could be explored and recommended if proven reliable.

²³⁹ Conversational technology is currently offered by [Google](#), [IBM's Watson](#), [Amazon's Lex](#), [Microsoft](#) and [Facebook](#).

²⁴⁰ Science and Technology Committee, House of Commons, '*Robotics and artificial intelligence report: Fifth Report of Session 2016–17*, September 2016.

Differential Privacy adds mathematical noise to a small sample of the individual's usage pattern to obscure an individual's identity without statistically harming the general pattern. *Today, a large number of statistical analyses can already be done in a differentially private manner by adding little noise.*

8. By actively providing datasets and/or alternatively a strong framework for the creation and maintenance of datasets, the government could address the above discussed issues of privacy, bias and data sovereignty. Especially in sensitive areas (such as healthcare and banking) but also in areas of public interest (such as transport and journalism) a framework for *the creation and the maintenance of data can help to avoid problems of bias and privacy that would undermine public trust.*

9. Identifying areas of data interest. The government's policy paper '7. *Data - unlocking the power of data in the UK economy and improving public confidence in its use*²⁴¹ is a step in the right direction. To push this agenda further and to ensure that the UK remains at the forefront of data innovation, the government needs to actively identify areas and ways in which *the sharing of datasets should be encouraged.* EU-funded accelerators such as *digitalhealth* currently assist private businesses by providing in-depth knowledge of the NHS. This should become a two-way street especially in areas of public interest, with businesses to be encouraged to share data and knowledge. *It is important to create incentives for businesses to open and share their datasets.*

10. AI is a programming paradigm that is performing knowledge tasks and therefore takes part in the creation of knowledge. A general condition for this new knowledge is the availability of data. In the interest of UK citizens and businesses, *the availability of data should be stimulated in some areas and ensured in others.*

Mercedes Bunz

1 September 2017

²⁴¹ Department for Digital, Culture, Media & Sport, 7. Data - unlocking the power of data in the UK economy and improving public confidence in its use, March 2017.

Eur. Ing. David Burden and Professor Maggi Savin-Baden – Written evidence (AIC0061)

SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

Call for Evidence

This document has been prepared by Eur. Ing. David Burden of Daden Limited, and Professor Maggi Savin-Baden of Worcester University. David has been active in the field of AI (particularly chatbots and virtual characters) for over 10 years, and has authored and co-authored a number of academic papers on the area, and was a finalist in the British Computer Society's Machine Intelligence competition in 2009. Maggi and David have collaborated on a number of AI research projects, including a "covert" version of the Turing Test in which 40 students in groups of 3-4 had up to 3 one hour discussions sessions, with each group seeded with an undeclared chatbot "virtual student". Not one student during the experiment identified, or even raised a concern, that one of the "students" may actually be a computer²⁴². Maggi and David have also written on the topic of digital immortality, and are currently writing a book on Virtual Humans for CRC Press in the USA, part of the Taylor & Francis Group. This submission is being written in their personal capacity.

David Burden

Maggi Savin-Baden

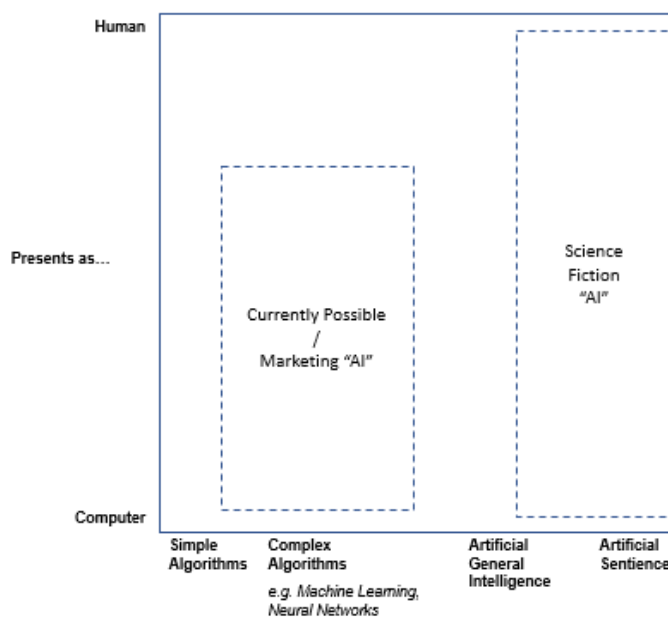
The pace of technological change

- *What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?*
- Before answering any of these questions we need to be clear about what is meant by the term "artificial intelligence". In public perception AI usually means science-fiction characters such as the Hal 9000 computer from 2001, or the androids from Channel 4's Humans. In modern marketing terms AI seems to be taken as describing almost any reasonably complex programme or algorithm – often based on machine-learning principles (e.g. " How Walmart is Using Machine Learning AI, IoT and Big Data to

²⁴²<https://www.researchgate.net/publication/309691977> Covert Implementations of the Turing Test A More Level Playing Field

Boost Retail Performance²⁴³ or “How AI is Impacting Content Marketing²⁴⁴). There has been a traditional observation that “what we call ‘artificial intelligence’ is basically what computers can’t do yet²⁴⁵ – once they achieve it we no longer think of it as needing “real intelligence”. With the current marketing hype the situation almost seems to have reversed into “‘artificial intelligence’ is basically anything that a computer can do now”.

- To help situate the current analysis Figure 1 represents the sophistication of the entity on one axis, and the extent to which it is presenting as human on the other, and how the current “marketing” interpretation of AI, and the public perception through science fiction map on to it. There is a clear gulf between what is typically described (especially by the media and marketers) as “AI”, and what the public perception of AI is as derived from science fiction. This often leads people to attribute more “intelligence” to systems than they actually possess – especially given the way that people typically anthropomorphise technology (e.g. think of Siri or Alexa as a person).



²⁴³ <https://www.forbes.com/sites/bernardmarr/2017/08/29/how-walmart-is-using-machine-learning-ai-iot-and-big-data-to-boost-retail-performance/#3c7fb20f6cb1>

²⁴⁴ <http://www.econtentmag.com/Articles/News/News-Feature/How-AI-is-Impacting-Content-Marketing-119095.htm>

²⁴⁵ <https://selfawarepatterns.com/2014/02/27/artificial-intelligence-is-what-we-can-do-that-computers-cant-yet/>

Figure 1: The AI Landscape

- In the future development of AI, and to bring the reality closer to the public perception, there are 3 main challenges; making virtual characters that look, sound and react in a human way, creating an artificial general intelligence – a general purpose problem solving machine, and perhaps ultimately endowing the AI with self-awareness. These are shown in Figure 2.

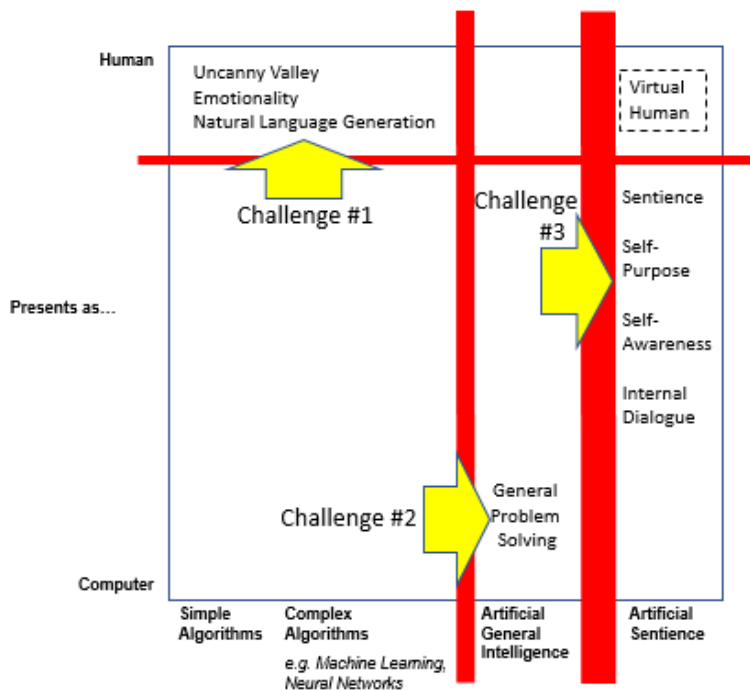


Figure 2: The Big AI Challenges

- In the recent development of "AI" there have been considerable improvements in the ability to present as human – such as better **text-to-speech, improved speech recognition and high quality avatars (Challenge 1)**. All these are driven not by AI research *per se* but by related fields such as Computer Generated Imagery (CGI) in movies and voice interfaces for mobile phones. The challenge in this area is though still to cross the "uncanny valley", the idea that human replicas may elicit feelings of eeriness in looks, sound and especially behaviour (e.g. emotional responses) from something that is almost human to something that can be readily mistaken for human.
- Where there has been seen less progress in real AI terms is along the

complexity axis. Whilst machine learning has significantly advanced and generated real benefits it shows little of being able to bridge the gap from just being a “smart” algorithm to being an **“artificial general intelligence” (AGI) (Challenge 2)** – a programme which can use common sense and self-directed learning to deal with a wide range of problems. This is probably the biggest AI challenge of the coming decade(s). The step from an AGI to something that is **truly conscious or sentient (Challenge 3)** is then probably of an even greater order of magnitude – if it can be achieved at all.

- There are undoubtedly social factors which will also effect this development, which are explored below.
 - *Is the current level of excitement which surrounds artificial intelligence warranted?*
 - The excitement is driven by a number of developments which are actually quite separate, but which are often taken together as representing AI. These are:
 - Virtual assistants such as Siri and Alex which provide **voice and conversational interfaces** to information and begin to deliver on some of the promise of virtual personal assistants and characters such as Hal.
 - The growth in the use of **machine learning techniques** to mine large amounts of data and to make deductions from it that can equal (or even exceed) human analysis.
 - The rise in the **level of autonomy** being given to computer controlled systems (which may incorporate machine learning or other techniques, or even conversational interfaces). Sheridan’s model of levels of autonomy is a useful reference here²⁴⁶.
- 1.1. Thus whilst the excitement promotes innovation the reality is that if “real AI” is seen as being represented by a system which is at least an artificial general intelligence then in reality development is likely to still be some 20 – 50 years away.

Impact on society

- *How can the general public best be prepared for more widespread use of*

²⁴⁶ Sheridan, T. B., & Verplank, W. L. (1978) Human and computer control of undersea teleoperators. (1978). Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab. A summary and critique of the scale is at <http://humanrobotinteraction.org/autonomy/>

artificial intelligence? In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

- The main issues that appear to be those of: Trust, Security and Realism.
- **Trust** – it is clear from our research that AI can be used effectively for teaching students but also persuade them to reveal more than they realise. For example, we used pedagogical agents to examine disclosure in educational settings. The study explored how the use of pedagogical agents might affect students' truthfulness and disclosure by asking them to respond to a lifestyle choices survey delivered by a web-based pedagogical agent. Findings suggested that users may feel comfortable disclosing more sensitive information to pedagogical agents than to interviewers, and that emotional connection with pedagogical agents was intrinsic to the user's sense of trust and therefore likely to affect levels of truthfulness and engagement. These findings support the growing body of literature which suggests that the social environment of cyberspace is characterised by more open, straightforward and candid interpersonal communication, and that a pedagogical agent can support this. AI, of whatever sort has social implications and raises questions what counts as privacy? Is it acceptable to use a webcam to spy on the person who is baby-sitting your child or film your roommate at university as a joke?
- **Security** - Enforcing security now carries a huge financial cost and therefore the protection of some areas are favoured over others. The question is what should be left unsecure and what might this mean? Furthermore, it is also important to consider whether there is now really any possibility of privacy or secure identities. It seems our identities can be both stolen or borrowed, as well as even used against us, perhaps not imminently but certainly in our future lives and work
- **Realism** – studies that compared human interviewees with virtual world chatbots (pedagogical agents in non-learning situations), that chatbots and human interviewees were equally successful in collecting information about their participants' real life backgrounds. To date, research into the realism of chatbots has formed both the greater part and the basis of use much of these technologies. Perhaps what we are really beginning to deal with here is 'augmented existence'. This notion of augmented existence is the idea that it is not just the tagging and integration that is affecting our lives but the fact that the meta systems themselves become a new means of categorisation.

Public perception

5. *Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?*

- There needs to be a way to bridge the gap between the public perceptions of AI – as already mentioned, the reality of what is currently being called AI, the very real impacts that even these marketing “AI” may have, and the potential impact if/when we ever get to a true AI. Whilst the last may be something that won’t happen for several decades when it does the consequences could be immense, and even the lesser “AI” technologies also have the ability to significantly change employment and offer significant challenges to moral and ethical norms.
- The opportunity is there for a clear and informed debate, making these clarifications, drawing on the popular interest in AI and science fiction and drawing out the possible road-maps forward and into the future, and the challenges which each bring.
- There are already events such as the BCS Machine Intelligence Competition, the Loebner Prize (an annual Turing Test implementation) and several “X-prize” type competitions which could be leverage to inform this debate.

Industry

6. *What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?*

6.1 Education could benefit by the use of intelligent tutors which are already being trailed in the US and to some extent in the UK and which could be used across education more extensively. The evidence suggests that pedagogic agents, virtual tutors and virtual mentors have a role to play in supporting student and staff education, enabling 24/7 learning in a personalised way and at a rate and in a style suited to the learner, and providing access to personal (and confidential) support. For example, one recent development was the creation of an automated teacher at the University of Edinburgh²⁴⁷ in the context of a Massive Open Online Course (MOOC). Chatbots have also been used in primary and secondary education children and young people about ethics, values and society. Virtual tutors typically draw on either the ability to support a “flipped classroom” model – so the human teacher can focus on discussion, not imparting facts, or the anonymous intimacy model to help overcome student anxieties.

²⁴⁷ Bayne, S. (2015) ‘Teacherbot: interventions in automated teaching’, *Teaching in Higher Education*, 20 (4), 455-467, DOI: 10.1080/13562517.2015.1020783.

6.2 An alternative form of virtual tutor is the AI driven non-player-character within a game – which is seen as becoming increasingly central as an informal learning approach by young people (e.g. Ark Survival²⁴⁸), that can teach them team work, problem-solving and strategy skills. We may ultimately see the merging of the formal and informal learning virtual tutor.

6.3 The anonymous intimacy effect could also be used by police and social services to engage with potential criminals in chat rooms, as well as using them with victims who may not speak to a human being but are prepared to speak to a bot.

6.4 There are also a range of potential combat and non-combat roles which AI could fulfil in the military²⁴⁹, such as those echoing the civilian roles described above. The more combat orientated roles would obviously need due consideration of the associated moral and ethical issues.

Ethics

8. *What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.*

6.5 There are ethical concerns where AI agents are designed to achieve specific purposes such as supporting online shopping, promoting services or goods, or supporting student learning. There are two key ethical issues at hand here. Firstly, there is the issue of whether research participants are aware of whom is 'behind the computer' – do they believe that they are engaging with a human, or with an artificially intelligent technology? This leads to the second ethical issue of participant willingness to disclose information to non-human researchers.

6.6 A major ethical consideration is the emergence of digital immortality²⁵⁰. Digital immortality is the continuation of an active or passive digital presence after death. Whilst society is beginning to grapple with "passive" digital immortality (e.g. through Facebook memorialisation) even current chatbot technology could create a far more active presence for the deceased, and with AGI and the creation of "cyber-twins" the distinction between a person physically alive or dead could cease for most of the people they interact with. Quite apart from family moral and relationship issues there are also ethical and legal issues that would need to be considered and protection against exploitation by companies seeking commercial gain.

²⁴⁸ <http://www.eurogamer.net/articles/2017-09-01-ark-survival-evolved-review>

²⁴⁹ <https://www.theguardian.com/uk-news/2014/mar/16/mod-secret-cyberwarfare-programme>

²⁵⁰ Savin-Baden, M. Burden, D and Taylor, H. (2017) The ethics and impact of digital immortality Special Issue of *Knowledge Cultures* - Technologies and time in the age of global neoliberal capitalism 5(2) 11-29.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

6.7 There are two perspectives that the Government needs to take – the first seeing AI as just the business as usual development of computer technology (which encompasses most of the current “marketing AI”), and the second looking AI at the eventual level of an AGI.

6.8 In the first perspective all the normal concerns about privacy, security and safety need to apply – although as with most current IT the more powerful the system the greater the potential impact of any breach.

6.9 Set against the 3 “developments” of AI discussed earlier, Government (and industry and the public) need in particular to be aware of:

6.9.1 With human-like and conversational interfaces the potential ethical issues that arise if a user thinks that they are talking to a human, or that the conversation is excessively engineered to encourage the user to reveal information (see the author’s paper on disclosure when talking to chatbots – the so-called anonymous-intimacy effect²⁵¹).

6.9.2 With machine learning the potential privacy issues of mining large amounts of data to reveal private information from public data.

6.9.3 With autonomy the levels of regulation that apply to systems operating at different levels of the Sheridan model, based on the risks of failure.

6.10 From a development and use perspective there is no reason for the Government to adopt any different measures than they do for any other form of ICT – these systems first and foremost need to be fit-for-purpose and deliver benefits over and above any alternative way of achieving the same goal.

6.11 It is in the area of artificial general intelligence and in crossing the uncanny valley that a more pro-active government stance in the development could be welcome. These are real challenges, which if overcome could have significant societal impact, and provide significant advantages to UK plc. However even here any investment should be focussed where it could have the most impact – and where other funding is less likely to be forthcoming. As already mentioned game and film CGI is already making the running in creating lifelike avatars so there seems to be little point in adding funding to

²⁵¹ Savin-Baden et al. “‘It’s Almost like Talking to a Person’: Student Disclosure to Pedagogical Agents in Sensitive Settings” in *International Journal of Mobile and Blended Learning* 5(2):78-93 · April 2013

that, but the pure challenge of AGI, and the specific challenges of synthetic argumentation or emotional/motivational models for AI would be more appropriate for funding.

6.12 Looking to the longer term the Government does need to start to consider the impacts of the higher levels of AI. As discussed in Question 8 there is the potential even now to start to create cyber-twins of real people, which could act on their behalf whilst their “host” is alive, and then continue to operate after their host’s death. The ethical and legal issues of the standing of such entities is something that may need discussion (and legislation) sooner rather than later.

6.13 There needs to be fast and in-depth research on the hidden impact of AI as well as its value. Current studies are small scale and local, funding is required for large-scale studies that investigate the potential impact and issues of AI across sectors (education, social, industry) in technical, social, ethical legal and regulatory terms.

4 September 2017

Michael Butterworth, Ms Joanna Goodman, Dr Paresh Kathrani, Dr Steven Cranfield and Chrissie Lightfoot – Written evidence (AIC0104)

Michael Butterworth, Ms Joanna Goodman, Dr Paresh Kathrani, Dr Steven Cranfield and Chrissie Lightfoot – Written evidence (AIC0104)

[Submission to be found under Ms Joanna Goodman](#)

Cancer Research UK – Written evidence (AIC0219)

Cancer Research UK response to the Select Committee on Artificial Intelligence

September 2017

1. Cancer Research UK (CRUK) is the world's largest independent cancer charity dedicated to saving lives through research. In 2016/17, we spent £432 million on research in institutes, hospitals and universities across the UK. Our vision is to accelerate progress so that three in four people survive their cancer for 10 years or more by 2034.
2. We are currently exploring the potential of artificial intelligence (AI) in cancer diagnostics, treatment and research, as well as the role that medical research charities should play in this space. We welcome this committee's call for evidence and would like to ensure that the potential for the role of AI in healthcare and medical research is recognised.
3. The UK is well-placed to be a world leader in the development and implementation of new health technologies such as AI; thanks to its capabilities in medical research and computer science, and the unique benefit of the NHS. We are therefore pleased to see that the Government has recognised the significance of robotics and AI via the Industrial Strategy Challenge Fund²⁵². We were also pleased to see that the potential of AI was referenced in Professor Sir John Bell's Life Sciences Industrial Strategy²⁵³.
4. We see significant potential in the use of AI in both medical research and care. We have identified specific opportunities in the use of AI in the early detection and diagnosis of cancer, as well as in planning and optimising cancer treatments – although we know that the potential for AI reaches into many different areas of research and healthcare.
5. However, there are several considerations if we are to realise the potential of new technologies such as AI. These range from practical considerations such as ensuring adequate and secure data storage to ensuring good governance and strong public trust. There are also specific considerations for the use of AI in diagnostics, such as ensuring professional education can adapt to new technologies and updating workforce planning. This is an

²⁵² Innovate UK and Department for Business, Energy and Industrial Strategy (2017), <http://bit.ly/2q0he2d> (Accessed September 2017)

²⁵³ Office for Life Sciences (2017) Life sciences: industrial strategy. March 2017, Available at: <http://bit.ly/2w7Qd3G> (Accessed September 2017)

opportune moment for these discussions and we are pleased to see that this Select Committee has been convened.

Opportunities for AI research in the early detection of cancer

6. Cancer Research UK is in the early stages of exploring machine learning in our research. Machine learning can be defined as the application of AI, insofar as it enables computers to 'learn' without being programmed.
7. Early detection and diagnosis of cancer is crucial if we are to improve UK survival. International comparisons suggest that cancer survival in the UK lags behind other comparable countries²⁵⁴. A significant driver of this survival gap is because the UK is poor when it comes to diagnosing cancer early²⁵⁵. We have recently collaborated with the Turing Institute on a Data Study Group to undertake a week-long opportunity to investigate the potential for using machine learning and computational statistics algorithms, along with mammograms and background information, to predict who will go on to develop breast cancer and, potentially, define ways of determining who is most likely to benefit from being prescribed tamoxifen for the prevention of breast cancer, which is the most common cancer in the UK²⁵⁶.

Case study: Cancer Research UK Grand Challenge

Cancer Research UK is currently exploring the potential of AI in the early detection of cancer, by taking a machine learning approach to examine patterns of symptoms and behaviours within accessible datasets that could indicate the presence of cancer. These data sets may be medical (e.g. GP presentation patterns, prescription records, health insurance claims) or non-medical (e.g. social media activity, shopping history, online search history). There is an opportunity to employ deep-learning approaches to combine these data sets with other cancer risk factors, and devise methods to drive diagnostic investigation at an earlier stage and facilitate the early detection of cancer. Cancer Research UK's Grand Challenge programme seeks to explore this and is currently accepting expressions of interest from research teams¹.

²⁵⁴ Coleman M.P. et al (2011) Cancer survival in Australia, Canada, Denmark, Norway, Sweden and the UK, 1995-1997 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet*, January 2011, Vol. 377(9760), pp. 127-38. Available at: <http://bit.ly/2jouEWS> (Accessed September 2017)

²⁵⁵ *ibid.*

²⁵⁶ Cancer Research UK Cancer Stats, available at: <http://bit.ly/2wZqkjk> (Accessed September 2017)

8. A similar approach to the case study above could be taken to datasets including biomarkers, for example UK Biobank. The GRAIL programme is a strong example of this and combines genomic sequencing with data science in order to achieve earlier detection of cancer by focusing on circulating tumour DNA²⁵⁷.

Improving Diagnosis and Treatment

9. There has been research into the use of AI and computer aided image analysis within cancer diagnosis. For example a recent Nature paper describes a system to assess images of potential skin cancers, with similar performance to a dermatologist in the research setting²⁵⁸. It is important that further development and implementation of these approaches are based on robust evidence and that comprehensive usage can be achieved through appropriate investment in information technology.
10. Professional bodies representing relevant clinicians and health professionals are very important stakeholders to involve when considering the use of AI, particularly with regard to professional education, training and workforce planning. The Royal College of Radiologists and Royal College of Pathologists should be part of the dialogue when considering introduction of AI in common practice.
11. Cancer Research UK is funding a number of projects in the field of computer aided detection and image analysis. Many of these studies are based at UCL or the Institute of Cancer Research, including:
 - *A Computer-assisted 3D Navigation System for Endoscopic-Ultrasound-Guided Diagnosis and Minimally-invasive Treatment of Pancreatic Lesions: **Dr Dean Barratt & Dr Stephen Pereira***
 - *Application of inverse problems and modelling to the correction of artefacts in prostate diffusion MRI **Dr David Atkinson, Dr Alex Kirkham & Prof Simon Arridge***
 - *Developing MRI Magnetic Susceptibility-Based Cancer Oxygenation Mapping (SBCOM) and Investigating its Clinical Potential to Measure Hypoxia in Prostate Cancer (PCa) and Head and Neck Squamous Cell Carcinoma (HNSCC) **Dr Karin Shmueli & Dr Shonit Punwani***
 - *Automated image processing and treatment planning for real-time MRI guided radiation therapy **Prof Uwe Oelfke & Prof David Hawkes***
 - *Multi-parametric ultrasound imaging for assessment of tumour response to radiotherapy. **Dr Emma Harris***

²⁵⁷ GRAIL: <https://grail.com/science/> (Accessed September 2017)

²⁵⁸ Esteva, A. et al (2017) Dermatologist-level classification of skin cancer with deep neural networks, *Nature* February 2017, vol. 542, pp.115-118. Available at: <http://go.nature.com/2wZOXgB> (Accessed September 2017)

Case Study: Cancer Research UK funded 'Optimam' project

X-ray mammography is used for breast screening programmes worldwide and over two million women are screened in the UK each year. However, although screening achieves earlier detection leading to improved survival there are also associated harms. One harm is the number of women unnecessarily recalled for further assessment.

Currently there are about 100,000 recalls per year in England of which only about 15,000 result in cancer detection. In the UK concern has also been raised about over-diagnosis leading to the Marmot independent review into the benefits and harms of breast screening. The review concluded that for every life saved by screening there were three cases of over-diagnosis. Specifically there was concern about the natural history and treatment of the increasing number of cases of ductal carcinoma in situ (DCIS) found in screening.

'Optimam' aims to improve the detection and characterisation of breast cancer during screening and diagnosis. It also aims to reduce harm by optimising the use of new X-ray imaging technologies including: digital breast tomosynthesis and tissue characterisation by computerised image analysis.

During the project, researchers will evaluate and optimise imaging technology by virtual clinical trials.

Over the next five years key decisions will be made on how and whether to introduce new imaging technology into breast screening and assessment. The nationwide adoption of such enhanced imaging is expected to lead to a reduction in the number of unnecessary recalls from screening, earlier breast cancer detection and improved survival. Currently there is a lack of evidence on which to base strategic decisions about screening technology, which this project will address. Looking back at images with pathology and disease progression may help to reduce overdiagnosis and enable improved management of breast cancer.

12. Pathology is another element of cancer diagnosis which could be augmented with the use of more digitisation and artificial intelligence for tissue and image recognition to support automation and accuracy in pathology. This was touched upon by our recent report, *Testing Times to Come*²⁵⁹ and is also being explored further by industry²⁶⁰.

The development of new cancer treatments

²⁵⁹ Cancer Research UK (2016) *Testing Times to Come*, available at: <http://bit.ly/2oQUfsi> (Accessed September 2017)

²⁶⁰ Philips (2016) Philips enables digitisation of tumour tissue assessment to support UK pathologists to fight cancer, November 2016, available at: <http://bit.ly/2w6J83L> (Accessed September 2017)

13. There is also exciting potential in the use of these new technologies to develop and deliver cancer treatments. We are particularly interested in the use of machine learning to predict response to treatment, based on genetic data. This is currently under consideration for Cancer Research UK's Lung Matrix Trial group²⁶¹. This trial stratifies patient groups based on gene changes identified in their cancer cells. The different patient groups may then evolve throughout the trial as more potentially effective drugs are identified. A machine learning approach could then be taken to the stratification of these groups. There is precedent for this within other disease areas, for example machine learning has been used to predict how well patients will respond to drugs used in the treatment of depression²⁶².
14. We are also aware of further work to understand the potential for applying AI techniques in radiotherapy²⁶³. As well as enabling fast and precise treatment planning, these techniques could free up time for a stretched workforce to focus on other activities such as patient care, professional development and research.

Policy considerations for implementation of AI

15. The implementation of these techniques requires a skilled and ready workforce, which in turn requires appropriate education, re-training and ongoing learning resources. Staffing numbers should also be considered, as this has been seen to limit progress with genomics to date: the UK has significant skills gaps for key staff such as molecular pathologists, bioinformaticians, statisticians, clinical geneticists and genetic counsellors²⁶⁴.
16. The usage of machine learning techniques to reduce the burden on the workforce is welcome, for example in pathology and radiology. However, these techniques should be seen as a diagnostic tool rather than an indiscriminate replacement for a skilled workforce. Full rollout of such techniques may be many years away and so routine workforce planning should not be disrupted by the perceived potential of techniques that have not yet been fully developed.

Public perception

²⁶¹ Cancer Research UK, National Lung Matrix Trial, <http://bit.ly/2xYK2eT> (Accessed September 2017)

²⁶² PReDicT (Predicting Response to Depression Treatment)

²⁶³ Deepmind, Applying machine learning to radiotherapy planning for head & neck cancer. Available at: <http://bit.ly/2y4uBCT> (Accessed September 2017)

²⁶⁴ Association of the British Pharmaceutical Industry (2015) Bridging the skills gap in the biopharmaceutical industry, November 2017. Available at: <http://bit.ly/1Tk8b80> (Accessed September 2017)

17. Our past research with Macmillan Cancer Support has found that people with cancer are largely very supportive of their cancer data being used: for example, 94% of people with cancer supported their cancer data being used for research and 89% supported their data being used for direct care. However, support does not override a desire to be informed: 83% believed it is important that people with cancer are informed about the cancer registry.
18. This principle should also be applied to new ways of working with patient data. To ensure that this level of support keeps pace with the changing way that data is used, communication and public engagement is paramount. The work of the National Data Guardian and of Understanding Patient Data should be central to this. We welcome the work of Understanding Patient Data to develop common frameworks for speaking about data and understand that further work is ongoing to examine emerging technologies. We encourage the Committee to engage with this programme.

Data quality, access and storage

19. The quality of analysis performed by an algorithm is very much dependent on the quality of the data. This could be a significant barrier for the application of machine learning in healthcare: although NHS datasets cover the entirety of the UK, they are not always complete and require significant quality assurance within data-holding organisations such as Public Health England – and therefore there is sometimes a significant delay between the data being collected and becoming usable. If we are to realise the potential of machine learning in healthcare it is vital that this is given sufficient investment. Part of this involves ensuring there is sufficient administrative support within trusts to enter and quality assure the data.
20. This is a prescient issue for cancer. The cancer registry is a world-leading database, containing data on over 14 million historical tumours. Patient level data on chemotherapy and radiotherapy provision is also held by Public Health England, in extensive datasets²⁶⁵. However, there are barriers to the application of machine learning on such datasets. The cancer registry is extremely complex, containing data from up to 19 different sources, and cancer treatment datasets are unfortunately still relatively patchy in their completeness (despite recent improvements). If we are to realise the potential of machine learning in healthcare we must invest in improving data quality – we can only trust the conclusions of AI if they are based on high quality data.

²⁶⁵ The Systemic Anti-Cancer Therapy (SACT) dataset and the Radiotherapy Dataset (RTDS), Public health England

21. A further challenge is how the conclusions of AI-driven analysis can be validated: without being able to see each stage of decision-making (the 'black box' nature of AI) it can be extremely difficult to judge two differing conclusions reached by two different algorithms. The extent to which this is a problem depends on the specific application: AI-driven pathology could be compared to human interpretation, for example, but in other more novel applications this would be more difficult. In some cases, researchers could compare predicted risk to patient-level outcomes data. However, there would naturally be a long delay before outcomes data could be used. One solution to this could be the use of synthetic datasets to test algorithms, which would have the added benefit of avoiding privacy concerns as they would not include real identifiable patient data.
22. We have also heard consistently from researchers that the progress of research has been limited by delays in accessing data held by national organisations such as NHS Digital and Public Health England. While we have seen some progress over recent years, this is still somewhat inconsistent. It is important that sufficient resource is given to these organisations to support data access as the applications of AI increase.
23. Finally, we would like to emphasise that the scope and potential of AI is extremely broad; even within cancer research. Each potential use of AI is fraught with its own issues and policy considerations and should be considered differently. These issues depend on the type, volume and size of datasets involved; the governance and workforce considerations will also naturally vary based on the application in question.

13 September 2017

Capco – Written evidence (AIC0071)

Capco RADAR submission for the House of Lords Select Committee call for evidence on AI

- 0.1 Capco is a financial services focussed business and technology consultancy which offers strategic, operational & technology solutions to help clients achieve positive results. It is actively transforming the future of finance to create a resilient market of transparency, trust and capital strength.
- 0.2 Capco RADAR is the technology research and development capability of the business and contains individuals with expertise in the areas of artificial intelligence (AI) and machine learning (ML). The following Capco RADAR response has been compiled through the lens of experience across banking, capital markets, and wealth management.
- 0.3 ML and AI are two terms used within this report and for the avoidance of doubt a definition is provided below:
 - AI is the broader concept of machines being able to carry out tasks in a way that we would consider “smart” and;
 - ML is a current application of AI based around the idea that we should really just be able to give machines access to data and let them learn for themselves.

The pace of technological change

- 1.0 **What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?**
- 1.1 ML is not a new concept. The first time the term was mentioned was by Arthur Samuel in 1959, in his paper titled: ‘Some Studies in Machine Learning Using the Game of Checkers’. In the paper, Samuel outlined work completed to show that a computer can be programmed so that it can learn to play a game of checkers better than the person who wrote the program [1]. A vast amount of the techniques that are used in ML and AI, are also based on algorithms researched in the 50s and 60s; such as:
 - k-nearest neighbours algorithms (k-NN, a non-parametric method used for classification and regression [2];
 - naive Bayes classifiers (NBC, family of simple probabilistic classifiers based on applying Bayes' theorem) [3];
 - support vector machines (SVM, supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis) [4]; and

- neural networks [5].
- 1.2 Even though these ideas and concepts were discussed almost 60 years ago, it was not until 2000 that the industry started exploring them. In 1969, Minsky and Papert wrote a book titled: 'Perceptrons: An Introduction to Computational Geometry' and commented that "neural network research slowed until computers achieved far greater processing power" [6]. This is widely accepted as true and in line with the concept of Moore's Law [7] where exponential computer processing power growth has led to exponential usage of AI and ML. Furthermore, traditional computers are now able to process vast amounts of data and train models in a far quicker time, making the use of ML not only an attractive proposition but commercially viable.
- 1.3 Another important factor was the emergence of the World Wide Web (the web) in the early 90s, which has been central to the development of the information age [8]. The web contributed to an unprecedented increase in data sharing and accessibility of cross-domain information, providing a wealth of opportunities; with new ways of processing required to make use of and understand the ever increasing data sets.
- 1.4 In the present day, large technology companies such as Google have actively encouraged open sourcing, where their code is shared and large numbers of people can collectively benefit from their research and actively build upon in. Such corporate strategies foster open source communities, leading to applicable code and tools to be made readily available, allowing the easy development of enhanced algorithms and predictive models. As online open source communities mature so will the emergence of increasingly powerful trained algorithms; with an increasing ability to handle larger and larger data sets, generating highly accurate results and further increasing the ability to handle even larger data. The cyclical nature of big data, trained algorithms, and increased efficiency presents a strong indication of the predicted exponential utilisation of AI and ML in everyday life over the next 20 years [9]:
- 1.5 **Current state:**
- Basic ML capabilities have been explored; however, the machine intelligence era and AI is still far away.
- 1.6 **In 5 years:**
- Perfect translation models;
 - Enhanced and highly accurate natural language dialogue engines will be mainstream (*i.e.* ChatBots);
 - Generation of newspaper articles by AI will be more common.
- 1.7 **In 10 years**
- Autonomous vehicles will be widely used;
 - Household electronics will be voice operated;
 - AI machines that will be capable of abstract thinking.

1.8 **In 20 years**

- Non-biological intelligence will be a billion times more capable than biological intelligence;
- We will multiply our intelligence a billion-fold by linking wirelessly from our neocortex to a synthetic neocortex in the cloud.

2.0 **Is the current level of excitement which surrounds artificial intelligence warranted?**

- 2.1 Over the last two centuries increasing use and advances in technology have affected our lives and shaped our work and social environment. Profound changes have occurred at the hands of human invention; from increasingly complex factory machinery in the 1800s, to commercially available cars in the 1900s, and the internet and smartphones in the 2000s. However, the evolution of AI technologies has the potential to have an even greater impact than the industrial and digital revolutions combined.
- 2.2 The excitement around AI is soaring high, so much so that there is almost not a single day without news on innovative tools, new applications, and more importantly a lot of hype and buzz on how amazing the future may look. This excitement has been building over the past few decades and has really taken off in recent years with renewed interest as necessary infrastructures become more feasible and affordable.
- 2.3 At the same time, however, there is much concern and apprehension expressed about the risks of AI to society at large. Such anxiety is not new and historically has been experienced by many new technologies in the past century; however, the level of concern with AI has the potential to be amplified as it is poorly understood and derives from a familiar and perhaps disconcerting source for many - human natural intelligence.
- 2.4 Current technologies utilise advanced ML algorithms and give the impression of AI; however, true human intelligence replicated by a machine is currently a long way off. Accurate natural language dialogue engines (*i.e.* ChatBots) are able to recognise written or spoken text and automatically devise a suitable response; however, they are currently prone to error and irrelevant responses. This type of technology relies on a set of standardised responses that can be selectively improved over time to 'learn' a more suitable response. Whilst this is impressive, it is far away from true AI. In relation to finance, these type of ChatBots are already utilised by banks where customers can talk to their accounts with personalised advice, metrics, and saving tips.
- 2.5 Looking to the future, AI advances will help in the world of finance by allowing for complex analysis of big data, consisting of transactional history, frequency of payments, and more complex interactions. The end result of AI exploring big

data from banks will allow for the early detection of fraud, highly personalised financial products, and individualised smart savings.

- 2.6 Looking around, it could be said that AI has already changed many aspects of our lives and there are many other transformations yet to come. The pace of change is so fast that quantifying the effect of AI on society is better made on a year-by-year basis rather than measuring the technological advances in decades. In the not too distant future it is likely machines with true AI, that can engage in common-sense reasoning, attain knowledge in multiple domains, feel, express, and understand emotions, will have already begun their 'evolution' and will co-exist with humans in the long-term.

Impact on society

3.0 How can the general public best be prepared for more widespread use of artificial intelligence?

- 3.1 Subjects of AI and ML should be part of the school curriculum. This will enable citizens at a young age to better understand the benefits, challenges, and potentials of this type of technology. AI is driving an increase in automation and robots are replacing humans to perform a variety of tasks. It is therefore important that future employees are well prepared for the workplace of the future; with sufficient knowledge, skills, and adaptability to fast-pace technology change.
- 3.2 With education, it is important to demonstrate what the current uses of AI are. Many think that AI is a threat, however few are aware that it is something that is already used in their everyday life in order to enhance their user experience and satisfaction. Simple use cases that people can understand and relate to can shed some light and allow people to familiarise themselves with the subject. Such use cases might include:
- Netflix movies recommendation: Netflix suggesting movies based on the previous movies you watched;
 - Amazon: suggesting products that are more likely to be interesting to you, based on previous purchases or your similarities with other shoppers;
 - Facebook image processing: When uploading a picture on Facebook, it detects faces, making it easy to tag your friends.
- 3.3 Additionally, as AI is becoming more integrated in the business world, companies will be in demand of people with expertise in AI in order to develop their own capabilities and stay competitive. There will be a growing need for people with a strong background in mathematics, statistics, and ML to work on innovation, as well as software engineers for the implementation of future products, tools, and services.
- 3.4 When it comes to data ownership, it is a commonly held view that individuals should have the right to know how their personal data might be used. Using data

for ML models, means that the generated models encapsulate part of your data in future code used to perform subsequent predictions.

- 3.5 Three types of scenarios with regard to personal data and AI are envisaged: (i) where the owners of the data agree to permanently 'donate' data for the creation of an AI capability; (ii) data owners can withdraw their data but allow for the 'learned' code to persist within a model; or (iii) at any given point an owner can choose to remove from both, an existing database and all data within subsequent models that have utilised personal data through AI. The third scenario would require retraining of the models without the personal data which would be deleted. This process, if feasible at all, is likely to be computationally expensive and it is likely such a response to a request would be subject to batch processing on a rolling daily or monthly basis. New laws might be required to take into account AI systems that make use of such data and in all cases care should be taken to protect data and ensure it is not reverse engineered.
- 4.0 **Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?**
- 4.1 Founder and Executive Chairman of the World Economic Forum, Klaus Schwab believes the world is on the cusp of the 'Fourth Industrial Revolution', caused by the rise of robots and AI [10], with estimates that these technologies could potentially put 5 million people out of work by 2020 [11]. A joint study by Citi and Oxford University in 2017 estimates up to 77% of jobs in China could be at risk of automation and 57% of jobs across the Organisation for Economic Co-operation and Development (OECD) region [12].
- 4.2 In financial services, there are numerous opportunities to improve existing services, as well as create new markets and provide access to financial products for underserved markets. In terms of improving existing processes, ChatBots (using natural language processing) could save \$8bn (£6bn) in costs every year across global business by 2022, analysts at Juniper Research forecast [13]. In addition to saving banks money, these can be used to provide better quality of service to a customer and benefit them by saving time on hold to a call centre, and aiding those unable to travel to a branch. For example, challenger bank Monzo have recently deployed a ML algorithm to assist their customer care staff to solve customers' problems more quickly and efficiently. Furthermore, AI can also be used to better detect money laundering [14].
- 4.3 A trend is emerging of big banks partnering with technology start-ups and this is happening more frequently in the financial industry. A combination of entrepreneurship and a high-tech concentration of skills creates the perfect environment for innovation and bodes well for the rise of the AI industry in high-tech centred areas such as Silicon Roundabout around Old Street, London.

- 4.4 The theory of disruption states, amongst other things, to beware of new emergent players offering services to a previously underserved market, luring them away. There are many examples of this in the AI industry, where new markets can be served at reduced costs due to lower overheads, ultimately benefiting the consumer. Whether banks benefit from this or not, AI implementation serves two positive objectives: (i) high-skilled technical job creation; and (ii) improving either knowledge of activity or product access.

Public perception

- 5.0 **Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?**

- 5.1 Customers are currently enjoying the benefits of AI. When all customer service lines are busy, an online ChatBot can be utilised to quickly and efficiently resolve an issue. Such applications are already in place for checking your bank account balance, unblocking your credit card, or notifying your bank that you will be travelling.
- 5.2 Google Translate, with its Deep Learning models help consumers understand a piece of text written in another language. A traffic system, integrated with the Global Positioning System (GPS), can inform a driver about traffic and provide alternative routes based on historical journeys and predicted traffic flow for a particular time, day, and location. Uber optimises its waiting time so that customers wait the minimum possible for their driver. Image recognition systems added at airports help reduce waiting time for passport checks, with the new electronic passport gates.
- 5.3 To assist in engagement, it is important to help people to understand the technology. This should be initiated at a young age, with computer science classes that focus on an understanding of how human computations can be augmented with technology such as AI and ML. Resources are available online to assist parents teaching their children to code; however, this should also be the responsibility of the creators of curricula. It will not necessarily be important to know how to code, but understanding coding will help people in the future be more adaptable and flexible which will be important in a rapidly changing society. Coding itself will become more abstracted as it has since the invention of computers (from machine code to programming languages; to functional languages and beyond). Additionally, there are already early forays into developing AI specifically to write code [15].
- 5.4 More teachers should be equipped with the expertise needed to teach such skills, incentives should be improved in the education field for subjects such as mathematics, computing, statistics, natural language processing and handling large data sets.

- 5.5 The media also has a responsibility to inform the public, which requires an understanding of the technology, its potential, and its limitations. For example, many news stories appeared recently that Facebook had shut down its AI bot because it had created its own language. In reality, the researchers realised they had not incentivised the bot to speak in English to aid their own understanding. This kind of misreporting breeds fear through lack of understanding, and could be seen as lazy headline-grabbing journalism.

Industry

6.0 **What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?**

- 6.1 In financial services there are a number of significant benefits that can be derived. There is the potential to work for under-served markets, whether that be in the UK or abroad as there will be lower overheads and the ability to offer the same level of service and advice to a greater number of people.
- 6.2 There will be the ability to increase personalisation with customers getting bespoke products that exactly meet their needs. It will be possible to detect fraud more easily and reduce risk. Financial services stand to benefit significantly as it is a service based industry with little manual labour required. Implementing AI into information technology (IT) systems and current processes will be much simpler and does not require the manufacture of expensive new bespoke equipment as may be required in other industries.

7.0 **How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?**

- 7.1 Although there is some influence of the amount and quality of data available in training AI which does lead to a potential of data-based monopolies, there is perhaps not as large an influence as expected. If you consider examples such as DeepMind's AlphaGo, it was not more data that allowed it to beat Lee Sedol but a smarter algorithm and more processing power. Although it looked at past games to learn how to play an important part of the development was playing itself and learning from that. When it comes to applications of AI however there may well be an advantage in certain industries where having more data will lead to an insurmountable advantage.
- 7.2 There is no natural way to address these monopolies and perhaps they should not be addressed while they serve the consumer and society, but it will require a vigilant regulator to ensure that companies do not abuse this to stifle competition.

Ethics

10.0 What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

10.1 Some points to consider for the financial services industry are:

- How do we monitor algorithms to ensure that they are not propagating historical discrimination of socio-economic status during decision-making (e.g. credit score, loan applications, fraudulent payments)?
- Who is accountable for an algorithms design? If a corporation uses an open-source library, how do we monitor and enforce accountability?

9.0 In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

9.1 No response submitted.

The role of the Government

10.0 What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

10.1 Financial services is already a highly-regulated industry, however, regulation rarely mandates use of a particular technology, it regulates outcomes. The UK Financial Services and Markets Act (FSMA) outlines objectives which are market confidence, financial stability, consumer protection and reduction in financial crime. These are guided by principles of good regulation [16].

10.2 AI can be used to detect fraudulent payment which makes regulation objectives more robust, but many financiers do not believe that current regulators have a deep enough understanding of the technology to successfully regulate [17].

10.3 As an example of the dangers of the technology, an algorithm trained on data that contains historical bias will maintain that bias. A recent paper looked at the way algorithms can be optimised for fairness by shifting the cost of poor classification from disadvantaged groups to the decision maker [18].

10.4 The General Data Protection Regulation (GDPR) will be applicable in the UK from May 2018 and mandates protection of individual rights against the risk that a potentially damaging decision is taken without human intervention. A definition of processors and controllers in the regulation will need to specifically look at AI solutions in reference to accountability of algorithm designs and implementations.

- 10.5 Regulators will need to put in place a risk framework and evaluation criteria for AI-based products and solutions. This will need to be done in conjunction with the foremost researchers in the field and an understanding of the creation of algorithms and the data set used to train them.
- 10.6 An existing challenge is that technological implementation to exploit a method of value capture moves much quicker than the regulatory implementation to rebalance regulatory objectives. One should also bear in mind that regulation of business processes places a proportionately heavier burden on smaller companies than it does large, which could be counter-productive to increasing competition for consumers.

APPENDIX

Abbreviations

ML	Machine learning
AI	Artificial intelligence
FS	Financial services
k-NN	k-nearest neighbours algorithms
GDPR	General Data Protection Regulation
FSMA	Financial Services and Markets Act
OECD	Organisation for Economic Co-operation and Development
IT	Information technology
GPS	Global positioning system

References

1. Samuel, A.L., *Some studies in machine learning using the game of checkers*. IBM J. Res. Dev., 1959. **3**(3): p. 210-229.
2. Larose, D.T., *K-nearest neighbor algorithm*. *Discovering Knowledge in Data: An Introduction to Data Mining*, 2005: p. 90-106.
3. Murphy, K.P., *Naive bayes classifiers*. University of British Columbia, 2006.
4. Hearst, M.A., et al., *Support vector machines*. *IEEE Intelligent Systems and their applications*, 1998. **13**(4): p. 18-28.
5. Yegnanarayana, B., *Artificial neural networks*. 2009: PHI Learning Pvt. Ltd.
6. Mynski, M. and S. Papert, *Perceptrons: An introduction to Computational Geometry*. MA: MIT Press, Cambridge, 1969.
7. Moore, G., *Moore's Law: Made Real by Intel Innovation*. Santa Clara, Calif: Intel, 1965.

8. Cooley, R., B. Mobasher, and J. Srivastava. *Web mining: Information and pattern discovery on the world wide web*. in *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on*. 1997. IEEE.
9. Makridakis, S., *The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms*. Futures, 2017.
10. Schwab, K., *The fourth industrial revolution*. 2017: Crown Business.
11. Caruso, L., *Digital innovation and the fourth industrial revolution: epochal social changes?* AI & SOCIETY, 2017: p. 1-14.
12. Frey, C.B. and M.A. Osborne, *The future of employment: how susceptible are jobs to computerisation?* Technological Forecasting and Social Change, 2017. **114**: p. 254-280.
13. Research, J., *Chatbots: Retail, eCommerce, Banking & Healthcare 2017-2022*. Juniper, 2017.
14. Kingdon, J., *AI fights money laundering*. IEEE Intelligent Systems, 2004. **19**(3): p. 87-89.
15. Balog, M., et al., *Deepcoder: Learning to write programs*. arXiv preprint arXiv:1611.01989, 2016.
16. Force, B.R.T., *Principles of good regulation*. London, Cabinet Office, Regulatory Impact Unit, 2003.
17. Cooper, R., et al., *The Role of Government in the Regulation of Financial Markets–New Rules for New Players*. 2017.
18. Hardt, M., E. Price, and N. Srebro. *Equality of opportunity in supervised learning*. in *Advances in Neural Information Processing Systems*. 2016.

Submission from:

Capco RADAR
www.capco.com

Authors:

Jibrán Ahmed (corresponding author)
Christos Aniftos
Sara Feenan
Stephen Harrison
Danushka Jayasinghe
Charles Laing
Jaspal Puri

Capco – Written evidence (AIC0071)

Jai Rajyaguru
Shabnam Rashtchi
Brian Vargas-Meinl

4 September 2017

CBI – Written evidence (AIC0114)

Introduction

1. The CBI welcomes the opportunity to respond to the Select Committee's call for evidence on the Artificial Intelligence (AI). We are the UK's leading business organisation, speaking for some 190,000 businesses that together employ around a third of the private sector workforce. Our membership is made up of businesses of all sizes, sectors and regions.
2. Digital innovations are at the heart of economic, social and cultural development across the UK. They drive productivity, help to raise living standards and lay the foundations for tomorrow's world. When businesses embrace data-sharing, innovation and digital technologies they create more jobs, generate investment and boost exports. By embracing digital to make new products and technologies, businesses across the UK – from healthcare to manufacturing - can make our lives better.
3. Digital is a major driver of business innovation, making access and adoption of technology essential for all firms and by extension, it is not an optional extra – for business and consumers. Data is the cornerstone of the UK's service based economy and this call for evidence is a window to marry the UK's strengths in digital with the opportunity of AI.
4. As set out in the CBI's 'Adopting the future²⁶⁶' paper - the world stands on the brink of technology-driven change. Every day new technologies are being developed, adapted and brought to market. This has resulted in fundamental shifts in the economy and created new challenges but has also thrown open doors to new economic opportunities. AI is one such technology that has captured the imagination of businesses. Our report highlighted that half of CBI members believe AI will fundamentally transform their industry. Clearly, AI offers economic opportunity but its transformative nature creates challenges for citizens and policymakers alike. The UK must be bold and businesses and government must work together to plot a way forward to develop solutions to the coming challenges.

²⁶⁶ CBI, Adopting the Future: Digital Adoption Survey, 2017:
http://www.cbi.org.uk/index.cfm/_api/render/file/?method=inline&fileID=37D3F2D7-B75D-4539-8F9399F1FD4C801F

5. There are several key measures the CBI advocates the government should consider when broadly shaping the future of the UK regulatory environment for AI:

- The UK Government should set up a joint Commission involving business, academics, employee representatives and Government to examine the impact of new technologies, including AI and robotics, on people and jobs, to develop recommendations for action and policy.
- The creation of the industrial strategy challenge funds offer an opportunity to coordinate government and industry to solve challenges with AI. The first challenges looking at the role of robotics and AI in extreme environments are one opportunity, but Innovate UK and the Department for Business, Energy and Industrial Strategy should explore other areas where AI could help. For example, cities could compete to implement the first driverless bus network or reduce hospital admissions to treat chronic diseases.
- The Department for Business, Energy and Industrial Strategy and the Department for Digital, Culture, Media and Sport should assess how to increase business uptake of readily available technologies and management practices, which will pave the way for firms adopting cutting-edge technologies such as AI in the future.
- The Department for Digital, Culture, Media and Sport, together with the Information Commissioner's Office (ICO) and wider stakeholders, should ensure the GDPR remains the chief legal framework that governs AI. Improved data-sharing can be achieved via guidance on contract clauses, technical leadership on APIs and opening-up of Government and public sector data sets. This will work towards enhancing trust and spur improved data sharing. ICO leadership, in cooperation with global partners, in how AI technologies can operate under the GDPR is to be encouraged.

1. The pace of technological change: Pioneering businesses are already adopting AI, but its use will increase over the next five to ten years

1.1. AI is likely to have a profound impact on business

1.1.1. The level of excitement around AI is warranted: businesses expect its impact to be profound. In response to our *Adopting the Future* survey, 2017 half (49%) of companies said that they expected AI would fundamentally transform their industry.

1.1.2. An overwhelming majority of companies expect AI will enhance efficiency (79%) and increase competitiveness (74%). Businesses also see AI playing a key role in improving customer satisfaction (72%) and differentiating them from their competitors (69%).

1.2. Pioneering firms are adopting AI now and over the next five to ten years, adoption rates are set to pick up further

1.2.1. In our *Adopting the future* survey, we asked firms to categorise their approach to adopting technology as pioneers “early adopters and champions of digital innovation” (34%), experimenters “curious about digital innovation and regularly experiment” (39%) or followers “wait until digital innovations are mainstream before trying them” (27%). Businesses’ likelihood to invest in AI was highly dependent on this categorisation.

Exhibit 1: Most businesses expect to invest in AI at some point, but followers risk getting left behind

Is it part of your business’s strategy to invest in AI/cognitive computing? If so over what time frame? (% respondents, by self-categorised approach to technology)

	Invested in the last 12 months	Plan to invest in the next 12 months	Plan to invest in the next 5 years	Plan to invest in 5-10 years	Do not plan to invest	Not familiar with the technology
All businesses	21	19	23	9	29	8
Pioneer	79	52	23	30	9	0
Experimenter	21	38	58	20	42	33
Follower	0	10	19	50	48	67

Source: CBI Adopting the Future Survey, 2017

1.2.2. Four fifths of pioneers said that they had invested in AI in the last twelve months, and half had plans to invest further in the next twelve months. But companies that identify as followers lag well behind. No followers have already invested in AI in the last twelve months and only 10% plan to invest in the next twelve months. Two thirds of these

companies describe themselves as being unfamiliar with the technology.

1.3. Challenges around adoption of AI vary by firm type, but skills are a key theme

- 1.3.1. For followers, the key challenges with adopting AI are highly related to limited levels of skills, knowledge and understanding. They cite the top barriers to adopting data-driven technologies as being a shortage of specialist skills (53%) the ability to identify a return on investment (45%) and internal understanding of AI (43%). These businesses often lack the history of investing in tried and tested technologies more generally, which can make a step towards AI even more challenging.
- 1.3.2. For businesses, more used to investing in cutting edge technologies, the challenges shift. For pioneers, skills are still the top challenge, cited by 41% of business, however, companies are less likely to report skills gaps at all levels of their organisation, suggesting they are more used to recruiting for digital skills. The second and third challenges are notably more focused on external issues: cyber security and privacy (36%) and the ability to raise investment capital (35%).
- 1.3.3. Just as the challenges for business are different, so too are the kinds of support they require from government.
 - **As a starting point The Department for Business, Energy and Industrial Strategy and the Department for Digital, Culture, Media and Sport should assess how to increase business uptake of readily available technologies and management practices, which will pave the way for firms adopting cutting-edge technologies such as AI in the future.**
 - **Industrial strategy challenge funds can be used to support pioneering businesses and create social value. The funds offer an opportunity to coordinate government and industry to solve challenges with AI. The first challenges looking at the role of robotics and AI in extreme environments are one opportunity, but Innovate UK and the Department for Business, Energy and Industrial Strategy should explore other areas where AI could help. For example, cities could compete to implement the first driverless bus network or reduce hospital admissions to treat chronic diseases.**

- There is a role for government to get the regulatory framework right as well, which is discussed later in this submission.

2. Industry: The use of AI will become increasingly widespread

2.1. AI has widespread application across sectors

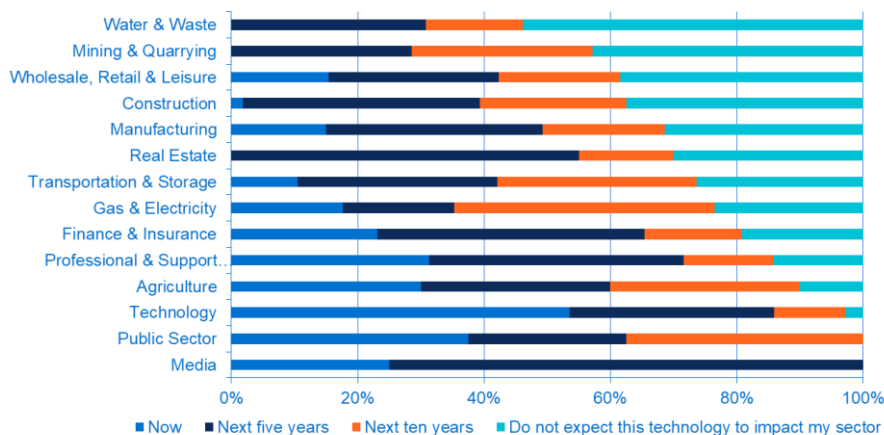
2.1.1. AI is applicable to a wide variety of industries, and its widespread use is only set to grow. The CBI's Innovation Survey 2016 showed that while around a quarter (23%) of companies thought that AI was currently impacting their sector, two fifths (37%) expected an impact on their sector in the next five years and a further one fifth (19%) expected an impact in the next ten years. The expected impact of AI is broader than for any other disruptive technology.

2.1.2. The companies that were most likely to report that AI was already impacting on their sector were focused in areas such as technology (54%) and professional services (31%). These sectors are the vanguard for the emergence of AI and will set the template for its delivery else through best practice or provision of AI technologies in themselves.

2.1.3. However, looking further out the impact is expected to be broader. In an increasingly digital future, it's not just the traditional early technology adopters that are looking at the role that machine intelligence might play. Of the construction firms that responded to the survey, 38 per cent believe AI will have an impact in the next five years.

Exhibit 2: The applications of AI are becoming more widespread

When do you expect AI to impact the sector in which your company operates? (% respondents)



Source: CBI Innovation Survey, 2016

3. The role of the Government, the impact on society and public perception: The growth of AI must be underpinned by a regulatory environment that facilitates data sharing and trust

3.1. Increased use of AI creates huge opportunities for businesses and consumers, but it will change the way data is used, and the UK regulatory environment needs to keep pace

3.1.1. Digital technologies provide new methods of engaging with consumers and businesses have been on the front foot to ensure that consumers trust that their data is being processed in a secure and safe manner. This includes pioneering new “just in time” push notifications on apps, which notify users if their data is about to be used in a way they might not expect.

3.1.2. But advanced data use is changing the very nature of business and creating increased complexity. AI will be the next step in advanced analytics and will capture both personal but also non-personal data, such as machine to machine generated data.

3.1.3. As data use changes, UK rules must ensure companies can compete in the global market. As such intervention into the data sharing market should reflect the diverse nature of AI's application. Addressing concerns around cyber security, privacy and ensuring a level of human control will be at the heart of AI development.

3.2. Like any other data-driven technology, AI has implications for privacy and data protection. GDPR should be the key legal mechanism to ensure AI development is tackled in an ethical and accountable manner

3.2.1. As AI technologies are increasingly deployed and interact with consumers, it will be important to ensure a high degree of trust and confidence in their use. Under the General Data Protection Regulation (GDPR), which will come apply directly from May 2018, consumers will have increased powers to access, correct and transfer data between providers. In addition, stricter rules will apply to the collection and use of personal data by organisations.

3.2.2. Aspects of the GDPR, such as consent, transparency and restrictions on processing sensitive data, will have implications for businesses looking to use AI:

- For example, the GDPR makes it clearer that an individual's consent must be "unambiguous" and that it must be a "clear affirmative action" such as ticking a box on a website. Furthermore, the data controller must be able to demonstrate that the consent was given, and the data subject must be able to withdraw that consent. Data controllers must also inform consumers specifically about "the existence of automated decision making including profiling and information concerning the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject" (Article 13(2)(f) of the GDPR)²⁶⁷.
- The GDPR includes a new reference to 'transparency'. Under the Regulation's lawfulness, fairness and transparency principle personal data must be processed lawfully, fairly, and in a transparent manner in relation to the data subject. This new express transparency principle will embed practices that help give

²⁶⁷ Official Journal of the European Union, REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016. http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

citizens increased control and importantly, a deeper understanding of how their data will be used by AI technologies.

- The GDPR prohibits the use of an individual’s sensitive personal data for automated decision-making purposes, unless: they have given their explicit consent or such automated decisions are necessary for reasons of public interest such as tax collection.

3.2.3. Increasingly, concerns about how the above can be reconciled with “blackboxing” have come to the fore. It is vital that efforts to boost transparency and accountability in this area focus on the ends and not the means. Proposed solutions such as the disclosure of raw data will be meaningless to the vast majority of users, while calls for the exposure of source code risk impinging on trade secrets and Intellectual property law, increase security risks and open up the AI to abuse and fraud. Alternatives exist to achieve the desired results, companies regularly provide FAQs sections to how AI functions or can audit the AI by inputting specific data to identify discrimination in the outputs.

3.2.4. Responses to this issue need to be proportionate. There is not a one-size-fits all approach to this issue. AI is deployed across a wide spectrum of businesses and is utilised in a variety of contexts, from automatically checking contractual compliance to access to credit and spam filtering. The use of AI in some contexts will require increased cooperation between businesses and regulators compared to other more routine measures.

3.2.5. The Information Commissioner’s Office has also analysed how data protection tools, such as anonymisation, PIAs and privacy by design, can help organisations ensure that AI complies with high standards of data protection. The ICO has welcomed industry’s efforts to build their own ethical principles and building relationships of trust with the public²⁶⁸.

3.2.6. Ultimately, is important to allow new data rules space to breathe and operate. The GDPR should be viewed as the key legal mechanism

²⁶⁸ ICO, Big data, artificial intelligence, machine learning and data protection, 2017. <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

ensuring AI development is tackled in an ethical and accountable manner.

3.3. The free flow of data across international borders will be vital to the success of AI

3.3.1. AI will be enabled by the global free flow of data. AI technologies will need to access, process and share data across multiple landscapes. Data localisation laws, such as mandatory storage of data in one jurisdiction or an inability to share data between different countries will inhibit growth and adoption. The UK is particularly at risk of data isolation if it unable to secure a mutually recognised adequacy agreement with the EU during the Brexit negotiations.

3.4. Data use is prevalent, a simple and flexible environment for data access and sharing is key

3.4.1. Data-use and collection are prevalent across every sector of the modern economy. The growth of so-called “data-monopolies” needs to be considered carefully in the context of the nature of data, the existence of data-sharing mechanisms and user behaviour.

Data is not finite

3.4.2. Data is sometimes referred to as the new oil but in reality, data is not a finite resource. Instead, data can be gathered, analysed, processed and then shared in to be repeated in a different method to create new value. Data has no intrinsic value in its raw sense but provides economic and social benefits when it is processed. The use of one data-set for a commercial benefit does not inhibit the use of the same data for social or non-commercial purposes.

A flexible and simple environment for data access and sharing can the delivery of AI’s economic and social promise and help deliver prosperity for all.

3.4.3. While there is a concern around the perception that some companies are creating “data monopolies”, the reality of the nature and use of data is that the expertise in processing and application of data that creates value for organizations. Many companies hold swaths of unused or underutilised data because they lack the skills and talent to understand and unlock its value.

3.4.4. At present, data-sharing takes place across a variety of sectors through a combination of mechanisms such as contractual arrangements, Application Programming Interfaces (API), open data and public private partnerships. Interference with contractual freedom and other methods would be unprecedented. Businesses who invest capital into the building, maintain and protecting datasets have a right to seek a return of investment if they carry the reinvestment costs.

Existing laws can be retooled or existing via other mechanisms clarified to ease regulatory burdens related to data sharing.

3.4.5. Flexibility is key due to radically diverse nature of businesses, IT systems and approaches to data. This is chiefly achieved currently achieved via contracts which offer flexibility for both buyers and suppliers. Contracts help bridge the needs of radically different businesses from across sectors and business models.

3.4.6. The introduction of new legal rights, such as 'data ownership' risk disrupting the UK's data value chain. Currently the legal concept of 'data ownership' does not exist under UK or EU law and has consistently raised concerns during previous consultations

3.4.7. There are steps the UK can take to boost data sharing across the board, such as nonbinding guidance based on existing legislation, coupled with contractual solutions. Increased transparency in data-sharing laws, via the new Data Protection Bill, may also help address perceived market issues and foster increased data-sharing.

3.4.8. Any moves to develop and increase uptake of APIs through technical guidance and best practice for companies and public administrations are to be welcomed. Such action should focus on transparency and reflect the current contractual environment in place.

3.4.9. The Department for Digital, Culture, Media and Sport, together with the Information Commissioner's Office (ICO) and wider stakeholders, should ensure the GDPR remains the chief legal framework that governs AI. Improved data-sharing can be achieved via guidance on contract clauses, technical leadership on APIs and opening-up of Government and public sector data

sets. This will work towards enhancing trust and spur improved data sharing. ICO leadership, in cooperation with global partners, in how AI technologies can operate under the GDPR is to be encouraged.

3.5. The role of Government will be critical to unlocking the social and economic benefits of AI

- 3.5.1. Government has a role to play in guiding the future of AI to ensure it delivers on its potential to benefit society and to solve critical problems previously believed unsolvable. Guiding AI will require that we grapple with complex questions of ethics, privacy, discrimination and employment. Industry, regulators and Governments can work together to mitigate risk, drive accountability and reap the benefits of AI.
- 3.5.2. Citizens are concerned about how we can manage the impact of new technology, in a world where the pace of change can mean some people feel left behind. This is an important message for business and policy-makers alike.
- 3.5.3. Public perceptions on AI centre on concerns over job destruction. However, our *Adopting the Future* survey suggests that companies expect the impact of AI on tasks to be fairly balanced, with 60% of respondents saying data-driven technologies would create new tasks and 64% saying they will replace tasks.
- 3.5.4. Positively, those companies with the most experience of investing in technology, were considerably more likely to think new technology would create tasks (87%), and – on balance – expect data driven technologies will create jobs.
- 3.5.5. While it is certain that AI, like any technology, will change the nature of work. The automation of repetitive and routine actions will create space for more creative roles or additional jobs that are created on the back of AI efficiencies. However, the impact on people will need to be carefully monitored and managed through collaboration between government, employers and employees.
- 3.5.6. **Businesses expect a large impact from AI and it is likely there will be social implications of its widespread use. To ensure the best possible social outcomes, the UK Government should set up a joint Commission involving business, academics, employee representatives and Government to examine the impact of new technologies, including AI and robotics, on people and jobs, to deliver recommendations for action and policy.**

6 September 2017

Center for Data Innovation – Written evidence (AIC0043)

Response to the Call for Evidence by the House of Lords Select Committee on Artificial Intelligence

On behalf of the Center for Data Innovation, we are pleased to submit the following response to the call for evidence by the Lords Select Committee on Artificial Intelligence. Paragraph numbers correspond to the question being answered.

The nonprofit, nonpartisan Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, DC and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as artificial intelligence, data analytics, and the Internet of Things. In our answers to the Committee’s questions, there are two particularly salient points we wish to emphasize.

First, there is little to no evidence to support the hyperbolic fears about AI, such as that the technology will cause cataclysmic job destruction, loss of privacy, bias and abuse, and even human extinction or enslavement. The notion that AI raises such grave concerns that policymakers should take a precautionary regulatory approach to limit the damages it could allegedly cause is both wrong and harmful to societal progress. However, there is substantial evidence of AI’s economic benefits. Thus, rather than attempt to limit AI, the role of policy should be to accelerate its development and adoption.

Second, over the long-term the potential benefits of AI are largely dependent on an adequate supply of data. Policymakers should therefore ensure they do not constrain the supply of data—such as by enacting overzealous data protection regulations—which would limit the positive impact of AI in jurisdictions where they apply, not to mention limit the growth of AI firms. Furthermore, policymakers should also work to close the “data divide”—the social and economic inequalities that may result from an insufficient collection or use of data about individuals or communities.

Yours faithfully
Daniel Castro
Director
Center for Data Innovation
Innovation

Nick Wallace
Senior Policy Analyst
Center for Data

1. “What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10, and 20 years? What factors, technical or societal, will accelerate this development?”

1.1 Current state of AI: AI is a field of computer science devoted to creating computing machines and systems that perform operations analogous to human learning and decision-making.²⁶⁹ The Association for the Advancement of Artificial Intelligence describes AI as “the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines.”²⁷⁰ AI does not necessarily imply machines with human-level intelligence, or machines that think in a human-like way. In fact, the very term “artificial intelligence” is a misnomer. Rather, AI describes a broad range of systems designed to behave in ways that humans think of as intelligent, and the level of intelligence in any given implementation of AI can vary greatly.

1.2 Contemporary AI systems generally exhibit one or more of the following functions: monitoring data to identify anomalies and patterns; extracting insights from large datasets in order to discover new connections and stimulate new solutions; predicting how trends are likely to develop; interpreting unstructured data that was hitherto difficult to classify; interacting with connected sensors and actuators in the physical environment; and interacting, communicating, and collaborating with humans and other machines. Practical applications of AI may involve just one or two of these functions, or may involve a complex array of algorithms performing near enough all of them, such as in autonomous vehicles.²⁷¹

1.3 Contributing factors: AI as a field of computer science began during the aftermath of the Second World War. Despite considerable excitement that major breakthroughs were just a few years away, the field showed only modest progress. However, the scientific and technological breakthroughs that have spurred its recent advancement, and made it more commercially viable during the last few years, are much more recent. Of particular importance is machine learning, a method whereby developers write algorithms that autonomously and iteratively build new analytical models in response to new data, without programming the solutions. Prior to this breakthrough, computer scientists had to laboriously pre-program outwardly intelligent behavior. The underlying factors that enabled the development of machine learning include better, cheaper computer hardware, particularly faster processing power and higher-capacity

²⁶⁹ Daniel Castro and Joshua New “The Promise of Artificial Intelligence” Center for Data Innovation, October 10, 2016. <https://www.datainnovation.org/2016/10/the-promise-of-artificial-intelligence/>.

²⁷⁰ “AI Overview: Broad Discussions of Artificial Intelligence,” AITopics, accessed September 29, 2016, <http://aitopics.org/topic/ai-overview>.

²⁷¹ Daniel Castro and Joshua New “The Promise of Artificial Intelligence”

storage, as well as a greater supply of machine-readable data and better algorithms.²⁷²

1.4 Development over the next 5-20 years: Machine learning will produce ever more advanced algorithms that can interpret and respond to more complex data in more sophisticated and more reliable ways. This will expand the variety and complexity of tasks to which computer scientists can dedicate AI tools in a reliable and commercially viable way. But contrary to speculation by some vocal critics of AI, the current progress of algorithmic development does not point towards the development of artificial consciousness, human-level or human-like artificial intelligence, sometimes called artificial general intelligence (AGI), anytime in the foreseeable future. Many of the dystopian fears about AI stem from the notion that AGI is imminent, feasible, or uncontrollable. In the 1960s, technologists began predicting that AGI was just a few years away. Since then, AI has progressed dramatically, and the underlying technology that supports it has developed even faster than predicted, yet AGI is likely just as far away today as it was 50 years ago. There is a very significant difference between the rapidly-advancing ability of machines to solve very specific problems in response to a narrow array of data supplied by humans, and a machine that can find solutions on its own to an infinite number of unpredictable and hitherto unknown problems with zero indication of what information might be pertinent to it.²⁷³ This difference is akin to that between a jet that can fly at the speed of sound and a spacecraft that travel at warp speed.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

3.1 The coming changes are not so dramatic as to require government to prepare the general public. The incoming wave of AI applications, though socially and economically important in its benefits, does not threaten to cause revolutionary social upheaval, especially not quickly. In fact, even the most socially consequential applications—such as the gradual emergence of autonomous vehicles—seems mundane in comparison to technological revolutions we have already seen, such as the rise of the Internet, not to mention the automobile itself, around which much of our urban infrastructure has been built. If policymakers act on the baseless assumption that AI has implications so dramatic as to require the public to be prepared, they risk creating undue panic, in turn generating political pressure for hasty policy decisions based on fear rather than fact and likely intended to slow down adoption. We have already seen such fears turn into ill-advised proposals to regulate and tax smart robots.

²⁷² Ibid.

²⁷³ Ibid.

3.2 The general public does not need special preparation, though the continued evolution of the workforce requires government to maintain and strengthen programs that offer job retraining and other supports for dislocated workers. AI will cause economic disruption in some sectors, but this disruption will come slower and affect fewer sectors than many popular commentators allege.²⁷⁴ For example, most doctors will not be replaced by AI, nor will nurses, journalists, civil servants, paramedics, or police officers. Taxi and bus drivers, airline pilots, and even lorry drivers will likely remain employable for the medium term due to remaining technological hurdles, consumer demand, public opinion, and public policy. However, light rail train drivers may face changes far sooner, as autonomous trains are already commercially viable and in use in urban subway systems around Europe.²⁷⁵ Policymakers should be prepared to help those who face such disruption retrain and find new career paths.

3.3 In most other fields, workers are more likely to find themselves working with AI than replaced by it. This will stimulate some demand for new skills, but the necessary experience will often be contingent upon industry-specific expertise that workers in that sector already have. Doctors, for example, will have to learn how to use some AI applications responsibly—but just as they will not be replaced by machines, nor will they be replaced by AI experts who are not doctors, and the AI tools they will use will have been designed with doctors in mind.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be limited?

4.1 As with most technology-driven efficiency gains, AI will benefit consumers and workers through increased productivity that will lead to greater choice and cheaper products and services and higher wages. This is particularly critical for the UK, which is suffering from an unprecedented productivity crisis, with productivity stagnant over the last decade. Unless Britain can find a way to boost productivity, social and political crises will increase as incomes stagnate, especially in the face of the increased proportion of retirees. AI will also lead to benefits for UK residents in a range of other areas, including healthcare, transportation, and environment. Those who stand to gain the least are people subject to types of social exclusion that restrict the supply of data pertinent to them, which in turn diminishes the relevance of AI tools to their circumstances.

²⁷⁴Ibid, Robert Atkinson and John Wu, "False Alarmism: Technological Disruption and the U.S. Labor Market, 1850-2015" Information Technology and Innovation Foundation, May 8, 2017. <https://itif.org/publications/2017/05/08/false-alarmism-technological-disruption-and-us-labor-market-1850-2015>

²⁷⁵ "One billion travelers on Europe's automated metro systems" Allianz pro Schiene, November 30, 2016. <https://www.allianz-pro-schiene.de/en/pressemitteilung/overview-automated-metro-systems-europe/>

4.2 Those gaining the most: Early uses of AI in healthcare are beginning to benefit patients, such as by helping doctors to identify problems in medical imaging and test results far earlier and more consistently than they might have otherwise.²⁷⁶ But the benefits of these tools remain as limited as health services' readiness or ability to deploy them—the technology, however, is out there already.

4.3 Businesses that invest in AI sooner will enjoy earlier rewards than those who turn to it later: for while AI will yield large returns for successful developers of it, like many other kinds of information technology, AI will help businesses in virtually all sectors to become more efficient and productive, and competitive markets will mean that workers and consumers will benefit. Citizens also benefit from early implementations of AI, such as autonomous vacuum cleaners, personal virtual assistants, and personalized language learning.²⁷⁷ AI is beginning to help lenders and insurers calculate risks more accurately using an unprecedented supply of data.²⁷⁸ This helps to accept or reject applications more wisely, and fine-tune premiums, interest rates, repayment periods, excess, and the quantity lent or value insured. For applicants and wider society, this promises to improve access to these financial services.

4.4 Those gaining the least: Because of the important impact on productivity growth from AI, virtually all UK residents will benefit. However, those who stand to gain the least are those living in a state of what one can call “data poverty.”²⁷⁹ These are social groups about whom little data is ever collected, which limits the extent to which data-driven services can be of use to them. These tend to be groups that are already marginalized in myriad other ways too, such as refugees. To give one example of the potential dangers of data poverty: we already know that some societal groups experience higher rates of certain diseases than others, for reasons that are in some cases fully understood by the medical profession, and in others less so.²⁸⁰ This means that a paucity of data on a given

²⁷⁶ Nick Wallace “5 Q’s for Matteo Carli, Chief Technology Officer and Founder of xbird” Center for Data Innovation, June 29th, 2017 <https://www.datainnovation.org/2017/06/5qs-for-matteo-carli-chief-technology-officer-and-founder-of-xbird/>

Nick Wallace “5 Q’s for Eyal Toledano, Chief Technology Officer at Zebra Medical Imaging” Center for Data Innovation, June 8th, 2017 <https://www.datainnovation.org/2017/06/5qs-for-eyal-toledano-co-founder-of-zebra-medical-vision/>

²⁷⁷ Nick Wallace “5 Q’s for Mait Müntel, Co-Founder of Lingvist” Center for Data Innovation, July 17, 2017. <https://www.datainnovation.org/2017/07/5qs-for-mait-muntel-co-founder-of-lingvist/>

²⁷⁸ Nick Wallace “5 Q’s for David Hand, Emeritus Professor at Imperial College, London” Center for Data Innovation, January 30, 2017. <https://www.datainnovation.org/2017/01/5-qs-for-david-hand-emeritus-professor-at-imperial-college-london/>

²⁷⁹ Daniel Castro “The Rise of Data Poverty in America” Center for Data Innovation, September 10, 2014. <https://www.datainnovation.org/2014/09/the-rise-of-data-poverty-in-america/>

²⁸⁰ Erick Forno and Juan C. Celedón, “Asthma and Ethnic Minorities: Socioeconomic Status and Beyond” *Current Opinion in Allergy and Clinical Immunology* 9(2):154-60, April 2009 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3920741/>

community limits the usefulness to those communities of AI tools intended to help tackle such problems.

4.5 What can policymakers do: The most important thing policy makers can do is to communicate a message around AI that highlights the progressive forces AI represent. Just as UK policy makers have supported technological change from the steam engine to the Internet, and not given in to the demands of Luddites, they need to do the same today. Exaggerations about the impact of AI have led to many harmful policy recommendations, particularly the claim that automation bolsters the case for a universal basic income (UBI). This claim begs the question, because it assumes, contrary to evidence, that high productivity from automation will cause joblessness. UBI would increase social exclusion and unemployment, and reduce living standards, because it is not time-limited, which distorts incentives.²⁸¹ The UK government also should not succumb to techno-panic by following the path of some who propose harmful policies like taxing AI, regulating smart robots, or significantly limiting data access on which so much AI depends.

The UK government should support public R&D into AI, to help the UK become a global leader in this emerging field. At the same time, it should ensure that the education system produces more data scientists and computer scientists with a understanding of AI.

4.6 Finally, the government should take steps to address data poverty. Data poverty is not usually an isolated problem, but a symptom of broader social exclusion. As data becomes more important in the economy, there is a real danger that the economic consequences of social exclusion could become more severe. Attempts to tackle social exclusion, therefore, must be combined with more ambitious approaches to the collection and use of data in public policy and public administration.²⁸²

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

Pellowski et al "A pandemic of the poor: social disadvantage and the U.S. HIV epidemic" *American Psychologist* 68(4):197-209 May-June 2013

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3700367/>

U.S. National Library of Medicine, "Why are some genetic conditions more common in particular ethnic groups?" *Genetics Home Reference*, <https://ghr.nlm.nih.gov/primer/inheritance/ethnicgroup> Last updated August 22 2017, Accessed August 23 2017

²⁸¹ Robert Atkinson, "Robots, Automation, and Jobs: A Primer for Policymakers" *Information Technology and Innovation Foundation*, May 8 2017.

<https://itif.org/publications/2017/05/08/robots-automation-and-jobs-primer-policymakers>

²⁸² Daniel Castro, "Europe Should Promote Data For Social Good" *Center for Data Innovation*, October 3rd, 2016. <https://www.datainnovation.org/2016/10/europe-should-promote-data-for-social-good/>

5.1 Public understanding, and even demand for, artificial intelligence, can help accelerate its adoption. Policymakers can facilitate this understanding by doing three things:

1. Inform themselves about what AI is and what it is not, and use this information to speak and argue more intelligently and more honestly in policy debates pertaining to AI.
2. Promote data skills throughout the education system, particularly as part of vocational and professional training in fields where data and AI are likely to play an important role, such as medicine.
3. Encourage the use of AI in public services and ensure out-of-date regulations do not become an unnecessary barrier. For example, UK medical regulations currently pose challenges for testing AI with patient data.²⁸³

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

7.1 There are no "data-based monopolies," and the winner does not take all. Data is non-rivalrous: customers who give their personal data to one company can provide it again to another. There are thousands of companies developing AI tools using large datasets. Accumulating personal data confers economic benefits on a company, but it does not automatically create a monopoly.²⁸⁴ However, policymakers can boost competition by encouraging the free flow of data. For example, the law should extend the data portability rights of personal data subjects to users of systems (such as cars) that generate non-personal data, allowing those users to share that raw machine data with third parties, such as insurance companies.

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

8.1 The ethical implications of AI are specific to the circumstances in which AI is deployed. For example, lethal autonomous weapons will demand a more robust ethical framework than autonomous vacuum cleaners.²⁸⁵ Moreover, many ethical dilemmas commonly associated AI are independent of the technology. For example, a popular question is what a self-driving car should do when forced to choose between equally lethal alternatives. This is not a dilemma caused by AI,

²⁸³ Nick Wallace "UK Medical Regulations Need an Update to Make Way for Medical AI" Center for Data Innovation, August 12, 2017. <https://www.datainnovation.org/2017/08/uk-regulations-need-an-update-to-make-way-for-medical-ai/>

²⁸⁴ Joe Kennedy "The Myth of the Data Monopoly: Why Antitrust Concerns About Data Are Overblown" Information Technology and Innovation Foundation, June 6, 2016. <https://itif.org/publications/2017/03/06/myth-data-monopoly-why-antitrust-concerns-about-data-are-overblown>

²⁸⁵ Daniel Castro "'Ban the Killer Robots' Movement Could Backfire" Computerworld, November 13, 2015. <https://www.computerworld.com/article/3005204/emerging-technology/ban-the-killer-robots-movement-could-backfire.html>

but by cars. Cars are dangerous machines that kill a staggering number of people with great frequency. AI will mitigate this dilemma by significantly reducing the number of accidents, but the fundamental implications of sitting inside a metal object and hurtling it forward at considerable speed remain the same.

8.2 Ethical concerns about explaining algorithmic decisions in human terms have led the EU to legislate for a “right to explanation” in the General Data Protection Regulation. Whether an individual has a right to have a decision explained depends on the decision, not the technology used to make it. The auditing of algorithms should be appropriate to the decisions they make, and not to a separate standard that applies solely to algorithms, such as that set out in the GDPR. Moreover, such approaches assume that human decision making is objective, transparent, and unbiased, something research has consistently shown is often wrong.

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

11.1 The Chinese government is striving for leadership in AI by pushing state-controlled businesses to invest in developing and implementing the technology. Countries that would compete with China, or that would prefer not to see it dominate AI development, should put in place strategies that both support AI research and identify opportunities to deploy AI in industry, without mirroring China’s mercantilism.²⁸⁶

11.2 Japan is a useful example in demonstrating how this might be achieved. The Japanese government has developed a roadmap for the commercialization of AI tools, which complements the AI development funding the government provides.²⁸⁷ Admittedly, Japan’s rapidly ageing population means the country has less to fear from claims of job destruction, and might be expected to take a more proactive approach to deploying AI in its industries. But as mentioned above, these claims are exaggerated, so the UK would do well to formulate a similar strategy that ties investment in AI research to social and economic gains—not least because the British government already sponsors AI research anyway.

11.3 The EU is more of a cautionary example: it has tried to regulate AI too early, imposing rules that address theoretical concerns without respect for

²⁸⁶ Joshua New, “How Governments Are Preparing for Artificial Intelligence” Center for Data Innovation, August 18, 2017. <https://www.datainnovation.org/2017/08/how-governments-are-preparing-for-artificial-intelligence/>

“Why China’s AI push is worrying” The Economist, July 27, 2017 <https://www.economist.com/news/leaders/21725561-state-controlled-corporations-are-developing-powerful-artificial-intelligence-why-chinas-ai-push>

²⁸⁷ Joshua New, “How Governments Are Preparing for Artificial Intelligence”

evidence. The aforementioned right to explanation will not guarantee accountability in algorithmic decisions, because it isolates individual decisions, making it harder to identify algorithmic bias, at the same time as imposing pointless costs on business. Statistical auditing is a more practical way to root-out bias in automated decisions.²⁸⁸ Furthermore, the European Parliament has endorsed a report that already calls for the regulation of robots and speculates wildly about their capabilities and risks.²⁸⁹ Just as over-regulation of biotechnology during the 1980s allowed the United States to take the lead, the new regulations threaten to have similar effects on AI, ceding leadership to other regions.

11.4. The World Economic Forum is also a poor example, as it too has largely succumbed to the “AI is out of control” narrative. Klaus Schwab, head of WEF, writes that “We stand on the brink of a technological revolution that will fundamentally alter the way we live, work, and relate to one another. In its scale, scope, and complexity, the transformation will be unlike anything humankind has experienced before.”²⁹⁰ As we highlighted above, there is simply no evidence for such hyperbolic claims. Viewing the development of AI in these overblown terms is virtually guaranteed to lead to bad policy.

1 September 2017

²⁸⁸ Nick Wallace, “EU’s Right to Explanation: A Harmful Restriction on Artificial Intelligence” TechZone360, January 25, 2017 <http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm>

²⁸⁹ “Civil Law Rules on Robotics - TEXTS ADOPTED - European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))” <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN>

²⁹⁰ Klaus Schwab, “The Fourth Industrial Revolution: what it means, how to respond” World Economic Forum, Jan 14, 2016 <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>

Centre for Health Economics University of York – Written evidence (AIC0242)

Written response on behalf of the Centre for Health Economics (CHE), University of York to questions posed by House of Lords Select Committee on Artificial Intelligence.

Points of clarification

Two terms are used in the questions raised that need clarification.

First it is our understanding that *Artificial Intelligence* (AI) is usually defined in respect of machines that perceive their environment and take actions in response to that perception. A reading of the transcripts of oral evidence to the Select Committee suggests however a much broader interpretation has been adopted. In that broader sense AI encompasses machines that assist in the interpretation of (medical) data, either autonomously or through following pre-specified instructions. We will henceforth use this definition and regard AI as a set of assistive technologies that relieve medical practitioners of the need to (unaided) interpret data such as diagnostic tests.

Second the term *Productivity* used in relation to the economy or the NHS has a number of different interpretations. From an economics perspective productivity is interpreted as the conversion of *inputs* into *outputs*. In order to make sense of multiple inputs and outputs these are aggregated usually by means of a weighted summation in monetary terms. Hence, dividing the total value of output of an economy (its GDP) by the total number of workers gives the productivity measure *output per worker*. This is the productivity of the workforce but does not account for how efficiently the economy uses other inputs. For the NHS, researchers at the Centre for Health Economics at the University of York have derived and utilised a measure of productivity in which the total value of NHS output is divided by the total cost of *all* inputs used and this is called *total factor productivity*. Henceforth when we refer to the productivity of the NHS we mean this measure.

Responses to questions

1. Does artificial intelligence offer a solution to any productivity challenges faced in the NHS? If so, how?

AI will facilitate the use of computer based inputs either as an alternative to, or in conjunction with, the time of doctors, nurses and other health care professionals. Hence, it implies a realignment of the inputs into the production of health care. Given that it is expected that AI will be less expensive than health carers' time this would be expected to result in an increase in productivity.

There are a number of caveats.

First, if AI is viewed as a substitute for carers' time then it will free up time that will be devoted to other health care activities. In this case the overall input cost will have increased and whether the additional output (health care) that is produced more than compensates for that depends on exactly how the freed-up time is used. So productivity could in fact decline. This would also be the case if the quality of care (its value) suffers because reliance on AI results in poorer outcomes and treatments. In short it is not guaranteed that AI increases productivity — that depends on how exactly it is implemented and whether in practice the health carers' time it makes available is deployed appropriately.

Second, the productivity *challenge* faced by the NHS is more substantial than simply requiring an increase in productivity. The NHS has shown increases in productivity that often exceed those achieved by the economy in general but the challenge it has been given is to increase that *rate of increase* and sustain it year on year. It is therefore important to distinguish between a step change in productivity and sustained productivity growth. AI may assist in the former but seems unlikely to have an enduring impact on the latter. Hence, AI may increase productivity without necessarily solving the productivity *challenge*.

2. How is the impact of AI on productivity likely to compare to the impact of previous digital technologies (such as personal computing) used in healthcare?

As set out in the preamble to these responses, the measurement of productivity requires the aggregation of inputs and outputs. As a consequence it is not typically possible to attribute changes in productivity to a single cause. We cannot ascertain the extent to which digital technologies have impacted on productivity and we will not be able to determine the separate effect of AI.

3. How could the impact of AI on productivity best be measured?

The standard measurement of productivity does not facilitate the separate measurement of the impact of AI. If AI is introduced selectively in some geographic areas, or for some selective types of health care, it may be possible to ascertain its effect by means of looking at differential changes in productivity. This would however be (a) uncertain and (b) take time following the adoption of AI to be achieved. In particular the initial effect of investment in AI may be to increase expenditure on inputs with no immediate effect on output — a decline in productivity. We are of the view that practically useful and timely measures of the impact of AI *on productivity* are unlikely to be possible because productivity is an aggregated measure. There is however a prior question which does not appear to have been addressed — how does AI impact on the quality of health care? If the introduction of AI increases the quality of health care at modest cost, then there is a reason to pursue its introduction without attempting to measure any impact on productivity. In this regard AI might be regarded as a different method of treatment and as with any other proposed new method the fundamental question is whether it should be adopted. There is a substantial body of knowledge— the measurement of cost-effectiveness — that can be, and in our view should be, applied to this question.

4. Is the relative lack of integrated digital records in the NHS an obstacle to the NHS being able to deploy AI on a scale? If so, what is the extent of this problem?

This would seem to be an operational question and one for which we do not have the relevant expertise to address.

5. Is there any action the Government take to help realise any potential benefit from AI in healthcare?

Organisations within the NHS are under great financial pressure. Developments like the adoption of AI are *investments* which require resources. Diverting resources away from front-line services is increasingly difficult when resources are limited and demand for services is increasing. As indicated under our response to 3. above a fundamental question is whether devoting resources to the adoption of AI is an appropriate use of those scarce resources and the Government could encourage and facilitate an investigation of that question. Beyond that if AI is found to be of substantial benefit to patients the Government needs to facilitate the financing of specific investments of this kind in the NHS.

Martin Chalkley, Professor of Health Economics, Lead on Health Policy
for and on behalf of the Centre for Health Economics, University of York

13 December 2017

Centre for Public Impact – Written evidence (AIC0173)

The role of government (Q10): *What role should the Government take in the development and use of AI in the UK? Should artificial intelligence be regulated? If so, how?*

The Centre for Public Impact's objective is to strengthen the effectiveness of governments around the world. We believe AI has the potential to drastically improve the quality of government outcomes. Accordingly, in parallel to exploring the UK government's role in regulating AI and in mitigating the adverse effects of AI in general, there is an urgent need to consider the benefits and risks concerning its use *in government*.

Our research (to be published imminently) shows that, similar to its effect on businesses and other organisations, AI has the potential to transform the way government operates. AI will allow policymakers to offer better public services which are more attuned to citizen's needs and achieve superior outcomes. AI will, for example, enable government to tailor interventions to an individual's circumstances. It can also improve the quality of decisions taken by, for example, judges, social workers and doctors and will allow government to see patterns which are not visible to the naked eye.

However, we believe that there are two main risks for government.

The first risk is the risk that comes with *doing nothing*. Failing to deploy AI in government because of a lack of technical expertise, an appropriate legal framework, talent or any other reason constitutes a serious risk. If government does not make use of this technology, it will be providing lower quality outcomes relative to what would be possible if it was deploying AI. This would risk severely undermining the legitimacy of government. In certain domains (such as helping jobseekers find employment) private providers may, over time, begin to offer AI-strengthened services superior to those offered by government.

The second risk is for government to *get AI wrong*. Policymakers may use AI but they could do so in ways which perpetuate existing biases and inequities, or are seen as abuses of government power.

While such ethical concerns permeate the functioning of government irrespective of the technology, the introduction of AI is certain to exacerbate these issues and bring forth its own, new challenges. If it is done the wrong way, citizens may well reject the use of AI in government. This might, in the extreme, lead to a moratorium on the use of the technology in government. Given the potential of AI in improving outcomes for citizens, this would be a huge wasted opportunity.

In light of the above we recommend the Select Committee give due consideration to the benefits and risks of AI in government.

To balance both the risks of doing nothing and of getting it wrong we recommend government start using AI immediately in order to build up its expertise and experience. However, we also recommend starting in policy areas where the downside risks to citizens are limited. Equally, government needs to adapt existing accountability mechanisms to ensure they adequately protect citizens where AI is being used to make decisions using their data or about them.

AI has the potential to transform the way the UK government delivers outcomes for its citizens. Avoiding the risks is, however, far from straightforward. We hope the Select Committee will include the role of AI in government in its considerations and would be delighted to provide further evidence on the issues raised above.

We will be publishing a working paper on the effects of AI in government shortly. We are also working with several governments around the world on creating a consensus on how to maximise the benefits and minimise the risks of AI in government. We would be delighted to brief the Select Committee in full on our findings and on the ongoing conversations we have with other governments.

6 September 2017

Centre for the Study of Existential Risk and the Leverhulme Centre for the Future of Intelligence – Written evidence (AIC0237)

**Centre for the Study of Existential Risk and the
Leverhulme Centre for the Future of Intelligence –
Written evidence (AIC0237)**

[Submission to be found under Leverhulme Centre for the Future of Intelligence](#)

Centre for the Study of Existential Risk and the Leverhulme Centre for the Future of Intelligence – Supplementary written evidence (AIC0239)

**Centre for the Study of Existential Risk and the
Leverhulme Centre for the Future of Intelligence –
Supplementary written evidence (AIC0239)**

[Submission to be found under Leverhulme Centre for the Future of Intelligence](#)

CENTURY Tech – Written evidence (AIC0084)

5th September 2017

**Submission on behalf of the organisation CENTURY Tech:
www.century.tech**

1. The pace of technological change: Is the current level of excitement which surrounds artificial intelligence warranted?

1.1. Current attitudes towards AI paint a future in which technology can solve many issues. This is certainly warranted but sometimes misplaced. The future as depicted in films such as *Ex Machina*, where artificially intelligent agents have an intelligence equivalent to that of humans - the so-called singularity, or 'general AI' - is far off. However, the future in which automated technologies augment and assist us in specific tasks - 'specific AI' - is already here. We make extensive use of such technology in both frivolous and serious contexts: to help entertain us with technology like NetFlix and Amazon, and also to help keep us safer, with AI technologies being used in security and defense systems. The benefit of these technologies is already apparent, and will only increase as technology advances further.

2. Impact on society: How can the general public best be prepared for more widespread use of artificial intelligence? In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

2.1. AI technology is designed to take large data sets, recognise patterns and to continually adapt based on new information. In education, this can be truly revolutionary. The amount of data that can be gathered and integrated into AI technology via online learning platforms is too great for a human to process. AI technology can assist teaching staff by analysing learning data and identifying actionable insights, which in turn enables more effective teaching. AI software can also personalise and adapt far more quickly and more effectively to gaps in knowledge and misunderstandings than humans can, allowing students to reap the benefits of AI technology directly. AI can quickly assess students' learning: what is holding them back, where their strengths and weaknesses are, where gaps in knowledge exist and how they can be remedied. We know this, because CENTURY Tech, a new AI learning platform for schools, is already doing this.

2.2. Unfortunately, many schools and colleges in the UK are not yet sufficiently equipped to benefit from advanced technology due to hardware and bandwidth limitations. As well as investments in hardware, additional training for the teaching profession is required to make the most of education technology of this nature. Teachers and senior leaders could benefit from a deeper understanding of AI technology and its potential impact. This should sit alongside better research into how best to effectively integrate this sort of technology into classrooms and into teaching.

2.3. At a broader level, there is a need for computer science experts (including developers and data scientists) in order to ensure AI in education can continue to develop and benefit from innovation. Including more comprehensive digital skills knowledge and training in the curriculum would be beneficial. Incentivising more people to take computer science related school and higher education courses would also benefit the market.

3. Industry: What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

3.1. Education, thus far, has not benefited from AI technology. Many other industries - banking, entertainment, retail, security, medicine, manufacturing, among others - have made extensive use of AI technology and seen huge advances in both efficiency and efficacy. Many believe that education still has much to gain from the field of technology.

3.2. Education is currently failing to meet the needs of all learners. Too often it is pupils from a lower socio-economic background who fail to achieve: at GCSE level, nearly 50% of children claiming free school meals achieve no passes above a D grade. This is suboptimal both from a societal and economic perspective.

3.3. AI can help to address educational underperformance, both by augmenting and supporting teaching staff, enabling more effective lesson planning and interventions, and by supporting students directly throughout the learning process. AI systems in education can provide tailored, adaptive learning experiences for students, allowing them to progress at their own pace, focusing on the areas of greatest weakness and building on the areas of greatest strength. AI systems can seamlessly rectify gaps in knowledge at the point they occur, not weeks down the

line when it can be too late. Technology can accurately and swiftly identify students at risk, allowing teaching staff to step in to support students.

4. Ethics: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

4.1. All AI companies face ethical considerations however, with the right vision, culture and safeguards in place, AI technology has the potential to be a powerful engine for good. In the instance of CENTURY Tech, our purpose is to improve learning outcomes by leveraging AI technology. Whether we are designing new features, creating content or making broader business decisions, the impact on students and teachers is always our first and final consideration. There may be cause for concern however when companies have a broader remit or more opaque purpose.

4.2. Security and privacy is a fundamental consideration of any data company. CENTURY takes these matters very seriously; we have self-certified with the DfE as compliant with their privacy standards for cloud services for schools, which means we could and would never share data for commercial reasons or with advertisers. Companies in other industries are not necessarily required to meet such stringent data safeguarding requirements. This is an area that may require greater transparency, public understanding or even legislation.

5. The role of the Government: What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

5.1. The government should support companies making use of AI technology to provide socially beneficial outcomes for the general public. The public sector is a less financially viable market than the private sector, and is under particular pressure due to the current programme of austerity, however this should not prevent innovative and cutting edge technologies from entering the market. To reduce the risk of this happening, it is useful to have grants and programmes available to companies working in this space and integrating AI technology. For example, the EDUCATE programme is of great benefit to CENTURY Tech and other Education Technology startups: access to experts as well as to assistance in building evaluative models is invaluable. Rolling out programmes like these more widely would be a positive step in the right direction.

5 September 2017

Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis, Dr Valentina Rita Scotti and Dr Aysegul Bugra – Written evidence (AIC0051)

Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis, Dr Valentina Rita Scotti and Dr Aysegul Bugra – Written evidence (AIC0051)

[Submission to be found under Dr Aysegul Bugra](#)

Charities Aid Foundation – Written evidence (AIC0042)

Charities Aid Foundation response to House of Lords Select Committee on Artificial Intelligence Call for Evidence

DATE: 1st September 2017

- 0.1 Charities Aid Foundation (“CAF”) is a leading international civil society organisation (CSO). Our mission is to motivate society to give ever more effectively and help transform lives and communities around the world. We work to stimulate philanthropy, social investment and the effective use of charitable funds by offering a range of specialist financial services to charities and donors, and through advocating for a favourable public policy environment.
- 0.2 CAF’s in-house think tank, Giving Thought, undertakes policy research and analysis to understand the macro trends affecting philanthropy and the work of charities. As part of that work we have been exploring the impact of disruptive technologies such as AI on the work of charities and the ways in which people are able to support them.
- 0.3 We have kept our submission firmly focused on the impact on charities and charity donors. Partly this reflects our particular expertise, but also we believe that there are potentially significant impacts in this area that few are currently thinking about, and which thus need to be highlighted.

1. The pace of technological change

- 1.1 In terms of the state of AI in the charity world, our sense is that there are small pockets of exciting innovation set against a backdrop of low levels of awareness, skills and understanding. This is not a situation unique to AI: though it is difficult to generalise about such a diverse sector, charities often struggle to adopt and adapt to new technologies due to lack of resources and skills. One of the key points we wish to make is that charities will need support from government, the tech industry and forward-thinking funders to develop the skills and resources needed to realise the potential of AI for social and environmental good. Given the huge potential that this technology holds in this context, the opportunity costs of failing to involve charities could be enormous.
- 1.2 It is generally recognised that two key factors in the accelerated development of AI recently are a huge increase in the amount and availability of data and the development of more sophisticated algorithms (such as deep learning algorithms) that can use this data to refine and improve their performance. This represents a challenge, as availability of data may present a significant barrier to the successful development of AI in a charitable context. The data on social and environmental needs that charities could use to refine and target their interventions is often locked up in siloes within government and the private sector; and where it is available it is not presented in a consistent, usable format. The adoption of open data

standards across the public sector and in the private sector (and within the charity sector itself) would thus be of huge benefit to charities.²⁹¹ Perhaps even more problematic is the data on outcomes, or the social impact of charitable interventions, which would allow us to measure and assess their efficacy and efficiency. This will be vital if we are to enable the application of AI to the allocation of philanthropic capital, and – as we shall see – this is likely to be an area of enormous growth in a future where there is likely to be a huge range of high-volume, low-value automated transactions, if we wish to harness some of this potential pool of capital for social good.²⁹²

1.3 Examples of AI already being applied in a social good context include:²⁹³

-AI social media analysis for suicide prevention: US tech startup Bark.us provides an AI product that monitors children’s social media activity across a wide range of platforms to detect early signs of suicidal behaviour.²⁹⁴

-AI Chatbots providing medical advice: Arthritis Research UK have partnered with Microsoft to pilot a service based on its Watson AI that can provide users with tailored information about the condition.²⁹⁵

-AI live translation: The Children’s Society has begun experimenting with using Microsoft’s AI-powered live translation tools to try to overcome language barriers in its work with young refugees and migrants in London.²⁹⁶

-Using AI to tackle poaching: The Lindbergh Foundation in the US has developed a programme called Air Shepherd, which uses unmanned aerial drones to patrol conservation areas and record footage. They have worked with a company called Neurala to apply deep learning algorithms to the data from these drones, with the aim of teaching them how to recognise poachers.²⁹⁷

-Using AI to analyse scientific research papers: Mark Zuckerberg and Priscilla Chan’s philanthropic venture the Chan-Zuckerberg Initiative (CZI) has purchased a startup called Meta, which has developed an AI that can help scientists navigate, read and prioritize the millions of academic papers in existence.²⁹⁸

²⁹¹ E.g. The UK government’s [Open Standards Principles](#).

²⁹² For more, see Davies, R (2016) [Artificial Intelligence and social impact measurement: how do we get a Google algorithm for philanthropy?](#), CAF *Giving Thought* blog, 24th October.

²⁹³ For more see Davies, R (2017) [5 Ways AI is Already Having an Impact on Charity](#), CAF *Giving Thought* blog, 2nd June.

²⁹⁴ Johnson, K (2017) [“Bark.us saves teens’ lives by using AI to analyze their online activity”](#), *Venturebeat.com*, 11th July

²⁹⁵ Weakley, K. (2017) [“Arthritis Research UK introduces AI-powered ‘virtual personal assistant’”](#), *Civil Society*, 24th March

²⁹⁶ Roach, J. (2016) [“Microsoft Translator erodes language barrier for in-person conversations”](#), *Microsoft blog*, 13th December.

²⁹⁷ Moon, M. (2017) [“Drones and AI help stop poaching in Africa”](#), *Engadget*, 21st May.

²⁹⁸ Wagner, K. (2017) [“Mark Zuckerberg’s philanthropy organization is acquiring a search and AI startup called Meta”](#). *CNBC.com*, 24th January.

- 1.4 In broad terms, AI technology is likely to affect charities in four key ways:
- i) **Creating new problems that charities will be called upon to address:** AI, like many technologies, will have unintended negative consequences, which charities will be relied upon to solve.
 - ii) **Developing new ways of addressing existing problems:** AI allows the analysis of data at an unprecedented scale and speed, which could suggest completely new ideas for solving social and environmental problems.
 - iii) **Offering new ways of working that utilise AI to support traditional charitable organisations:** AI could help to find efficiency savings in existing approaches or be used to ensure that organisations learn from their data on impact and improve.
 - iv) **Creating new governance structures and operating models for achieving social good:** AI could lead to ways of working which augment or even replace traditional charitable organisations entirely.
- We shall touch on all of these in this consultation response.

Impact on society

- 2.1 AI is almost certain to have an effect on charities by altering the nature of social and environmental needs; or creating entirely new ones. This could pose major challenges for charities in the future if they are not only asked to spread their already-stretched resources even thinner, but also find that they struggle to develop the technical knowledge and skills required to understand these new problems and find solutions to them.²⁹⁹
- 2.2 Examples of areas where AI might exacerbate existing issues or create new ones include:
- Filter bubbles:** We have already heard a lot about the potentially damaging effects that social media “filter bubbles” that result from algorithmic bias can do, by limiting people’s experience and trapping them in echo chambers in which they find their existing views and prejudices reinforced and amplified. The growing ubiquity of non-traditional interfaces (e.g. conversational interfaces such as Amazon’s Alexa or Microsoft’s Cortana, or augmented/virtual reality interfaces in the near future) means that this effect is likely to be heightened. As a growing proportion of our experience becomes mediated by these AI-driven interfaces, the danger is that they will seek to present us with choices and interaction based on existing preferences and thus will limit our experience even further (perhaps without us even realising it). This will create new challenges for charities in terms of things like heightened social isolation and decreased community cohesion. It may also make it harder for charities to engage with potential supporters, both because they might struggle to break through the filter of the AI interface to make the first contact and because it may become harder to create an

²⁹⁹ For a more detailed exploration of the new social issues that AI and other technologies could create see Davies, R (2017) [Future Imperfect: 10 new problems that technology will create and charities will have to deal with](#), CAF *Giving Thought* blog, 13th April.

emotional connection if people’s empathy for those outside their realm of experience becomes diminished.³⁰⁰

-Online influencing: The 2016 US Presidential Election exposed the extent to which it is now possible to apply machine learning software to data on online behaviour to reliably predict and manipulate the way people will react to information they are shown. Platforms like Facebook are allowing companies like Cambridge Analytica to access up to five thousand data points on every user,³⁰¹ which enables them to create profiles of individuals which can reliably predict not only preferences but reactions to new media. Such companies are able to take advantage of the learnings of behavioural economics, and in particular the work of Daniel Kahneman -that people are much more likely to react to information by relying on emotional reflex (System 1) rather than a dispassionate analysis of information (System 2) – and employ “behavioural microtargeting” to deliver thousands of variants of content which is optimised to influence individuals.³⁰²

-Algorithmic bias: A growing amount of attention is being paid to the ways in which algorithms can entrench existing biases in the data sets they operate on, and the adverse effects this can have on individuals and even entire demographic groups.³⁰³ Given that many charities exist to represent the most marginalised people in society by ensuring that they have a voice and are able to exercise their rights and access services, this sort of targeted bias (whether intentional or unintentional) is a real source of concern. Charities could play a role not only in dealing with the symptoms of this problem when it occurs (by supporting victims of algorithmic bias), but also in attempting to prevent it by working with technology companies and government to provide oversight of the use of AI and algorithmic process and ensure that the unintended consequences are minimised.

-The Future of Work: AI could play a part in the large-scale transformation of the workplace; including the replacement of many white-collar knowledge-based roles that were previously thought to be relatively safe from automation. This would have an enormous impact on the shape of society, as we move to a world in which the majority of people no longer work. Solutions such as the adoption of some form of Basic Income have been proposed as ways to meet this challenge. However, given the centrality of the notion of work to our concepts of value and identity there are likely to be significant

³⁰⁰ For more, see Davies (2016), “[Is technology making us care less about each other?](#)”, *CAF Giving Thought* blog, 6th July

³⁰¹ Cheshire, T (2016) “[Behind the scenes at Donald Trump's UK digital war room](#)” Sky News. Retrieved 18 August 2017

³⁰² It has also been suggested that Cambridge Analytica played a role in the UK EU membership referendum, e.g. Cadwalladr, C. (2017) “[The great British Brexit robbery: how our democracy was hijacked](#)”, *Observer*, 7th May, although the firm has contested this claim.

³⁰³ For more, see Pickering, A. (2017) “[Algorithm’s Gonna Get You: What the rise of algorithms means for philanthropy](#)”. *CAF Giving Thought* blog, 18th January.

unforeseen consequences, and charities will have to play a key role in addressing them.³⁰⁴

-Inequality: Widening inequality is already one of the defining issues of our age. AI could exacerbate the situation by concentrating wealth and power in the hands of an even smaller minority of people who own and control the technology and its applications. Many of the charities that currently focus on campaigning against the corrosive effects of inequality will need to broaden the scope of their activities to include this new technological inequality.

-Digital Exclusion: Charities are already starting to play a key role in ensuring that their beneficiaries are not left behind by the pace of technological change by helping them to develop skills and giving them the opportunities to make use of things like the internet in a safe environment. As technologies like AI develop and converge they are likely to become ever more ubiquitous and access to them may well become a basic right (as the UN declared access to broadband to be in 2016). Charities will thus need to ensure they are in a position to help their beneficiaries when it comes to accessing these new technologies.

3) Industry

- 3.2 AI will offer new ways of addressing many of the challenges that charities currently deal with, and hence could make them more effective and efficient. For instance, real-time analysis of big data could be used for preventative services (such as the suicide prevention initiative highlighted above). Or AI could be used to automate advice services and interactions with service users (as in the Alzheimer’s Research case already mentioned). These advice services would not only be lower-cost, but could be more effective than human-led services at getting people the information they require and also can be made available 24/7 so that people can access them whenever they need them.
- 3.3 Charities will face many of the challenges that we highlighted in the previous section facing wider society, as well as having to help others deal with them. For instance, charities may find themselves on the wrong end of algorithmic bias when it comes to things like insurance, banking services or regulation. Likewise, the future automation of the workplace will affect charities just like any other organisations (although there may be a positive impact if an increase in the number of people no longer working to earn money leads to a rise in volunteering).
- 3.4 AI could also have an enormous impact on the way that charities raise funds. It could be used for philanthropy advice services, encompassing things like which causes are most pressing, which interventions most effective and which methods of giving are most appropriate. This kind of advice is currently the preserve of the very wealthy, but AI could lower the cost

³⁰⁴ For More, see Davies (2017) “[Giving in a World Without Work? Automation, Universal Basic Income and the future of philanthropy](#)”, CAF *Giving Thought* blog, 11th January.

sufficiently to make it a mass-market service.³⁰⁵ If this were to happen, it would open up new opportunities for charities to reach donors, but would also present new challenges: for instance, if the algorithms determining selection of charities proved to be biased towards already well-known organisations or towards popular causes, this could make it even harder for low-profile charities and those working in unpopular cause areas to raise funds.

3.5 AI also opens up the possibility of entirely automating the process of allocating philanthropic capital, by matching areas of most immediate need with the most effective relevant interventions based on analysis of big data. In a future where the expansion of the Internet of Things means that there are likely to be vast numbers of high-frequency, minimal-value machine to machine transactions taking place, we may hope to use a proportion of the revenue generated for charitable purposes. Given that it will be impractical for a human to have oversight of all these micro-donations, the application of some form of “AI philanthropy” is the most likely solution.³⁰⁶ This represents a vast new pool of potential income for charities: one which will dictate new approaches to fundraising based on social impact measurement and put an onus on organisations to understand how the algorithms which determine how philanthropic funds are allocated actually work.

3.6 AI may also be one element of an existential threat to the very notion of a charitable organisation. The convergence of AI with blockchain technology opens up the possibility of creating AI Distributed Autonomous Organisations (AIDAOs):³⁰⁷ decentralised organisational structures in which networks of human and AI-controlled nodes are able to interact and work together towards shared goals, without the need for a centralised structure for decision making, logistics or asset ownership. These structures may lend themselves well to charitable purposes, as they could democratise the processes of distributing assets to those in need or campaigning for social change. If this happens, it will at raise fundamental questions about the role that traditional charities could still play (perhaps as expert nodes or “oracles”, or as curators of social issues for others), but may even supplant the idea of a formalised, centralised charitable organisation.³⁰⁸

The role of the Government

³⁰⁵ For more see Davies, R. (2017) “Robotic Alms: Is artificial intelligence the future of philanthropy advice?”, CAF *Giving Thought* blog, 22nd May.

³⁰⁶ For more see Davies, R. (2016) *Giving Unchained: Philanthropy and the Blockchain*. London: Charities Aid Foundation.

³⁰⁷ For all of CAF Giving Thought’s work on blockchain technology and charity, see <https://www.cafonline.org/about-us/publications/blockchain>

³⁰⁸ For more, see Davies, R. (2017) *Losing the Middle but Keeping the Heart: Blockchain, DAOs and the future decentralisation of charity*. London: Charities Aid Foundation.

- 4.1 We agree with the recommendation of the House of Commons Science and technology Committee in its 2016 report on “Robotics and Artificial Intelligence”, that an ongoing Commission on Artificial Intelligence be established; and crucially that this includes representation from the charity and NGO sector.³⁰⁹ These organisations represent many of the most marginalised groups and individuals in our society, so it is vital that they are able to speak up for them in the debate over the development of AI and also that they are able to highlight concerns about the potential impact on the wider work of charities.
- 4.2 Government can also play a role in ensuring that charities and their beneficiaries are able to harness the potential benefits of AI technology by providing funding and support to develop skills in the charity sector and for work which seeks to boost digital inclusion.

Learning from others

- 5.1 There are existing initiatives in other countries which seek to explore the risks posed by the development of AI, and in particular the role that philanthropic funders can play in trying to mitigate these risks. For example, the Open Philanthropy Project has a dedicated focus on “Potential Risks from Advanced Artificial Intelligence”, through which it gives grants to support research and work in this area.³¹⁰

Authored by:
Rhodri Davies
Programme Leader, Giving Thought
Giving Thought
Charities Aid Foundation

Adam Pickering
International Policy Manager,
Charities Aid Foundation

1 September 2017

³⁰⁹ <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>

³¹⁰ <http://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence>

Mr Thomas Cheney – Written evidence (AIC0098)

Thomas Cheney
PhD Candidate in Space Law
University of Sunderland

Summary:

1. This paper examines several issues regarding artificial intelligence. It starts by discussing the definitions of the different applicable terms relating to artificial intelligence; arguing that there is a valuable distinction to be made between artificial intelligence, artificial moral agents, and robots. It goes on to discuss the potential impact on society and potential ways of mitigating that. Focusing on the need to educate the public about artificial intelligence and provide them with the tools and skill sets necessary to adapt to the changing economic environment. As well as discussing the need to adapt the welfare system and ensure the continuing competitiveness of the British people and economy. Then it discusses the issue of accountability and responsibility for the decisions made by artificial intelligences, with particular regard for those capable of making 'life affecting decisions' such as autonomous weapons and autonomous vehicles. Finally, it discusses the activities of the European Union and the United States in this field and the value they can have for this committee.

Defining Artificial Intelligence:

2. Defining artificial intelligence is important yet challenging. Part of this is due to the constant 'moving of the goal posts' regarding what constitutes artificial intelligence. Artificial intelligence is generally regarded as futuristic; therefore, it cannot be currently in existence. Science fiction has not helped here, people expect *2001's* HAL 9000 or *Star Trek's* Commander Data, they get IBM's Watson or Apple's Siri. Another aspect of the difficulty of pinning down a definition is due to a confusion and conflation of differing terms. Robot and artificial intelligence are often used interchangeably³¹¹, when they are two separate concepts. Robots can have artificial intelligence but not all robots are artificially intelligent and not all AIs 'reside' in robots. There is also potentially a third term to be considered, 'Artificial Moral Agent' or AMA, an AMA goes beyond a simply being an 'autonomous intelligent' system to one that makes moral decisions. AMA refers to systems that are more than just excellent computers, but systems that actually 'think', that should therefore be responsible for their decisions.³¹²

³¹¹Such as in the recent piece in *The Times Magazine*: 'How Artificial Intelligence Will Change Your Life Sooner Than You Realise: A Handy Guide for Humans (Not Suitable for Robots)' *The Times Magazine* (London, 19 August 2017) 19

³¹²See Wendell Wallach and Colin Allen, *Moral Machines: Teaching Robots Right from Wrong* (Oxford University Press 2009) for a more detailed discussion

3. There is value in separating 'AIs' and 'AMAs'. The *Oxford English Dictionary* defines artificial intelligence as "the performance by computer systems of tasks normally requiring human intelligence, such as translation between languages."³¹³ This seems a useful working definition of an 'AI', which is effectively a tool, similar, if not more advanced to the systems millions of us are already familiar with using on a regular basis. However, an 'AMA' is more than just a programme executing commands, it takes actual decisions, it makes moral choices, even if it is not 'conscious' or 'sentient', it should also be 'responsible' and 'accountable' for the decisions it makes. It is worth noting, even with the criticism that it receives, that the 'Turing test' can be passed by an 'AI' that appears to a human to be 'thinking' even if the 'AI' itself does not actually think, therefore mere mimicry should not be sufficient to rule out an 'AI' being considered an 'AMA'.

4. In this paper, artificial intelligence will be used both as a general catchall to describe the concept, 'AI' will be used in the manner described above, as will 'AMA'. However, it should be clear that defining the relevant terms is a complex matter in and of itself, and that any definition needs to be flexible and adaptable enough to deal with developments in artificial intelligence.

Impact on society and Public Perception:

5. The development of artificial intelligence is going to have a significant impact on society, especially economically. It will enable greater economic efficiency but is likely to lead to significant reduction and disruption in employment opportunities, at least in the short to medium term. There are several things that should be considered to help prepare the general public for the development of artificial intelligence.

6. First and foremost is education. This is not only about educating the public about artificial intelligence whose understanding of the subject is largely based on portrayals in movies, television and books but preparing them for the artificial intelligence enabled and transformed economy. As with the industrial revolution of the 19th century there will be many who fear and/or oppose this new technology. Part of this will stem from the 'killer' or 'crazed' robot trope common in film, television and books, but it will also stem from the economic impact of artificial intelligence, especially if mass job losses become a reality. Industry and supporters of artificial intelligence need to mount a PR campaign to explain the benefits of this new technology and restore some of the wonder about the future. To some extent this is already being done by organizations such as the Future of Life Institute.

³¹³*Concise Oxford English Dictionary* (12th edn, 2011), 74

7. The other side of the education coin is the most important, and the one where the government can have the biggest direct impact. Children and young people need to be given the skill set that will enable them to survive, compete and thrive in the artificial intelligence enabled and transformed economy. STEM subjects need to be promoted from an early age, computer literacy needs to be considered as vital as reading and writing, coding and programming should be taught to all students from as young an age as is possible and practicable. The British education system needs to be training students for the future economy, this will most likely require a radical transformation of the education system but it is vital if British workers and the British economy is to remain competitive or even viable. Furthermore, there needs to be support for people to retrain, it needs to become easier for people to return to education, university or otherwise, part time or full so that they can transition to the new economy.

8. Second, there needs to be an examination of the suitability of the welfare system to manage the potential impact of artificial intelligence on the employment market. Even in it is assumed that artificial intelligence proves to be simply a 'creatively destructive' force, the transition period is likely to see a return to high levels of unemployment or underemployment. It will take people time to retrain and it will take time for new industries and jobs to come into existence. It is also sensible to consider the concept of a basic income, and even the 'robot tax' proposed by Bill Gates.³¹⁴

Accountability and Responsibility

9. There are a number of ethical, legal and policy issues regarding artificial intelligence. Accountability and responsibility are two of the major concerns, this is particularly true for those systems that will make 'life affecting decisions', most notably autonomous vehicles and autonomous weapons systems. Accountability and responsibility are key. Who takes responsibility for the artificial intelligence when it does wrong, either intentionally or otherwise? This is certainly something that could potentially scale depending upon the capabilities of the artificial intelligence itself, i.e. as we move closer to human level intelligence it becomes more reasonable for the artificial intelligence to assume more of the burden of responsibility and accountability. However, punishment needs to mean something in order for it to act as a deterrent. Most humans fear, or at least strive to avoid, incarceration and punishment for wrongdoing, an artificial intelligence may not, and an artificial intelligence may not even fear 'death' (i.e. being turned off). Though it should be easier to 'reform' an errant artificial intelligence, as it should be a 'simple' case of rewriting their software (though that could open its own ethical 'can of worms'...)

³¹⁴Richard Waters, 'Bill Gates Calls for Income Tax on Robots' *Financial Times* (19 February 2017) Accessed at: <https://www.ft.com/content/d04a89c2-f6c8-11e6-9516-2d969e0d3b65> Last Accessed: 4 September 2017

10. However, it does seem reasonable to assume that 'human level' artificial intelligences are sufficiently distant to not be a pressing policy or legal concern and in the meantime, there are workable solutions to the question of where and to whom to look for accountability and responsibility. For corporate owned artificial intelligences, the most obvious solution is to appoint a named responsible officer, somewhat like the role of Data Protection Officer, who would be responsible for ensuring that the company's assets are compliant with the law, ethical guidelines, and standards. An Ombudsman or Watchdog could also be created to monitor compliance and levy fines and provide additional guidance. For artificial intelligences owned by individuals the same approach to legal responsibility could be taken as it is for pets; i.e. you are legally responsible for the actions of your artificial intelligence. For more advanced artificial intelligence's, a licencing regime would also probably be sensible. As far as giving artificial intelligences' themselves a legal identity, the law has long had 'artificial legal persons', primarily in the form of corporations, however, this is probably best left for advanced, human like artificial intelligences.

11. The question of responsibility is particularly important for those artificial intelligences which may have to make decisions with harmful or even lethal consequences for humans. This applies most notably for autonomous vehicles ('driverless cars') and, of course, autonomous weapons systems ('killer robots'). This is a timely issue as autonomous vehicles are currently under development and could be on our roads during the current Parliament.³¹⁵ Autonomous weapons systems are also under development.³¹⁶ Therefore we need to discuss these issues now to ensure that we have time to implement an appropriate framework.

12. Regarding autonomous vehicles, beyond the 'trolley problem', there is the basic question of who is responsible for the 'decisions' made by the vehicle. Is the owner, the operator, the manufacturer or the software developer liable? Does it make a difference if the owner and the operator are one in the same or not? Regarding the 'trolley problem' and its associated variations, should there be customer choice in how to answer or should a single mandated answer? (And, perhaps in consideration of the interests of international trade this should be coordinated at an international level?)

13. Initially, it would be reasonable for the person actually using the autonomous vehicle to shoulder at least some of the responsibility for the autonomous vehicle, however as the technology matures the need for a 'manual override' should lessen and thus responsibility could be shifted to the owner and/or the

³¹⁵Driverless Cars Trail Set for UK Motorways in 2019' *BBC News* (24 April 2017) Accessed at: <http://www.bbc.co.uk/news/technology-39691540> Last Accessed: 4 September 2017

³¹⁶Ben Farmer, 'Prepare for rise of 'killer robots' says former defence chief' *The Telegraph* (27 August 2017) Accessed at: <http://www.telegraph.co.uk/news/2017/08/27/prepare-rise-killer-robots-says-former-defence-chief/> Last Accessed: 4 September 2017

manufacturer. Of course, the degree of responsibility should scale with the degree of control, and any unauthorised tampering with the vehicle, the software in particular, should shift liability to the person who did or requested the tampering. This is an area where guidelines are probably the way to go initially, the reality of operation of autonomous vehicles should help clarify matters. That said, the overarching principle should be on the protection of the ‘innocent bystander’, i.e. the person not using an autonomous vehicle.

14. Autonomous weapons systems, more popularly known as ‘killer robots’ are a more pressing concern. There is a ‘Campaign to Stop Killer Robots’³¹⁷ which is calling for a pre-emptive ban on these weapons.³¹⁸ There are several arguments against them, but the common argument is that they are simple incapable of being compatible with International Humanitarian Law. Again, part of the issue comes down to responsibility and accountability. A human soldier can be put on trial for war crimes, a robot cannot. If these weapons systems are developed and utilized there needs to be clear accountability. Fortunately, the armed forces are not unfamiliar with the principle of superiors being held responsible for the actions of those under their command, however the details will need to be addressed.

Learning from Others

15. Artificial intelligence is a topic under discussion around the world. The UK should absolutely take notice of what others are doing, in the United States and the European Union especially. The European Parliament recently published a report on civil laws for robots³¹⁹, which this subcommittee should certainly take into consideration. US President Obama also released a report in 2016 which this subcommittee should consider.³²⁰ There may also be cause for a global coordination effort via the United Nations, as was done at the beginning of the Space Age. There probably is not a need for a ‘Committee on the Peaceful Uses of Artificial Intelligence’ but an intergovernmental group of experts or some similar arrangement would be worth considering.

5 September 2017

³¹⁷<https://www.stopkillerrobots.org/>

³¹⁸Samuel Gibbs, ‘Elon Musk Leads 116 Experts Calling for Outright Ban of Killer Robots’ *The Guardian* (20 August 2017) Accessed at: <https://www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war> Last Accessed: 4 September 2017

³¹⁹European Parliament’s Committee on Legal Affairs (2016) *European Civil Law Rules in Robotics* Brussels, Belgium: Policy Department for ‘Citizens’ Rights and Constitutional Affairs’

³²⁰National Science and Technology Council Committee on Technology (2016) *Preparing for the Future of Artificial Intelligence*

Dr Esyin Chew – Written evidence (AIC0166)

What are the implications of artificial intelligence? In Love and War

A. Introduction

1. I am a computer scientist and educationist who work with robots that looks like human. Collaborating with the South East Asia National Human Rights Institutions³²¹ including Malaysia³²², I investigate the educational implications for embedding humanoid robot activist in educating child rights in ASEAN regions³²³. The robot activist has artificial intelligent (AI) capabilities such as assessment and feedback analytics.
2. As a Welsh Crucible alumni³²⁴, I extended the humanoid robotics research to healthcare sector with a Welsh hospital and an International Rehabilitation Centre. I am a pastor's wife who use humanoid robot to engage refugees' and autistic children in learning and teaching.
3. As a Fellow of the Higher Education Academy, I am a productive author, reviewer and editor for quality research publications; an academic trainer and a leader of advanced technology enhanced assessment and feedback for higher education in international horizon.
4. The following are my personal views grounded on the above experiences and expertise:

B. My Definition of Artificial Intelligence (AI)

1. Generally speaking, AI is perceived as intelligent technologies, from computers to robots that mimic human's intelligence and five senses for learning, analytical reasoning, decision making, real-life problem solving and companionship.
2. I would assert that the prospects and future of AI lies in the hand of robotics, naturally, a moving AI agent acts as an extension of mankind (not as a replacement); like a pair of angel's wings to human, or Doraemon's pocket attached to human.

C. The Impact of AI on Society

Question 3. How can the general public best be prepared for more widespread use of artificial intelligence?

1. There are clear disparities between those who accept or worship the 'wonder of AI' and those who against it, such as "How AI is used to

³²¹ <http://www.suhakam.org.my/regional-international/seanf/>

³²² <http://www.rightsgodigital.com/nao-robot-activist> ;

<http://www.rightsgodigital.com/2016/testimonials>

³²³ **Chew, E.**, Lee, P. H., DharmaratneA., Chen, B. W. & Raju, D. S. (2016) SUHAKAM - Going Digital with Monash, *The 3rd Asian Symposium on Human Rights Education*, 4-6 Aug 2016, Fukuoka, Japan.

³²⁴ <http://www.welshcrucible.org.uk/esyin-chew/> ; <http://www.welshcrucible.org.uk/2017-2/>

transform Google Translate³²⁵, “Robotics Public Private Partnership in Horizon 2020³²⁶, “EU spends millions building robot that makes pizza³²⁷; versus “AI: we are like children playing with a bomb³²⁸ and “AI could lead into third world war³²⁹. Both camps attract general public to perceive AI differently with a mixed feeling: in love and war. The complexity arises from the complex nature of AI. At instrumental level the idea of AI is intuitively simple, to put human’s brains and senses into machines. However, its social-economical and psychological implications are far more complex. The AI technological advances are pervasive but the education reform are fall behind. In last two centuries, both technology and education raced forward in the UK and US, generating rising living quality and massive economic expansion, however, ‘technology sprinted ahead of limping education in the last 30 years, leading to the recent upsurge in inequality³³⁰. There is an urge to raise the stagnation in the level of AI education across the UK. The key question is that how do we prepare the UK general public for the impact of AI, the disparities and the future AI leaders from the UK to shape the directions not into the devil’s wings but the angel’s?

2. Hence, I would recommend to commence the public awareness and readiness through the two strands of education and healthcare. By seeing the living robots and by experiencing the AI in daily life, these strands are the least threatening and have wider accessibility for general public at all age:
 - a. **Education:** The educational policy makers need to have a continued passionate in embedding robot tutor in day-to-day classroom for motivation, personalised assessment and feedback: from pre-school, primary education to higher education. The personalised assessment and feedback provided by humanoid robot are featured with the real time data analytics capabilities such as individualised academic performance feedback and student sentiment (emotional / satisfaction / happiness) analysis supported with educational psychological theories. As a professional educationist, I would argue that, in the next wave of learning

³²⁵ <https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html>

³²⁶ <https://ec.europa.eu/digital-single-market/en/robotics-public-private-partnership-horizon-2020>

³²⁷ <http://www.telegraph.co.uk/news/2017/08/11/waste-dough-eu-spends-millions-robot-makes-pizza/>

³²⁸ <https://www.theguardian.com/technology/2016/jun/12/nick-bostrom-artificial-intelligence-machine>

³²⁹ <https://www.theguardian.com/technology/2017/sep/04/elon-musk-ai-third-world-war-vladimir-putin>

³³⁰ Goldin, C & Katz, L. (2010) *The Race Between Education and Technology*, Harvard University Press. <http://www.hup.harvard.edu/catalog.php?content=reviews&isbn=9780674035300>

innovation no longer lies at e-learning or mobile learning but, a thoughtful integration of face-to-face learning with a walking AI agent, a humanoid robot tutor. However, the public preparedness need to be met. My research show that this innovation enable students to gain high level of motivation in learning engagement for futurists' perspectives after the intervention³³¹. This practical suggestions are as follows:

- i. The UK government to facilitate and support the Universities-Schools collaborations. Universities that have expertise in AI can partner with local schools to develop robotic tutor for various subjects and implement it for educational intervention.
- ii. To initiate national pilot interdisciplinary projects supported by industries and corporates for Corporate Social Responsibilities. These can be carried out with selected or volunteered schools and universities for the robot tutor intervention.
- iii. To open industry-university research grant calls to support the AI and educational action research by key research and industrial funders. This is to accelerate the commercialisation and creativity of the robot tutors across educational sector.
- iv. To establish the National Institution for AI and Robotics in Education as a catalyst for excellence for national and international show cases.

There are some international examples to be referenced from such as National Human Rights Institution Malaysia teach human right to schools using a humanoid robot³³², Robots are helping out in Singapore pre-school³³³, Robot tutor in Japan to teach English³³⁴ and a step forward in using robotic tutors in primary school classrooms in Spain³³⁵.

- b. **Healthcare:** Telemedicine and AI in healthcare are increasingly pervasive in the international spectrum. Labour is costly and intensive for healthcare in the UK. There are language, communication and attitude barriers between patients and medical teams, and among medical teams. A personalised medical

³³¹ Chua, X.N. & **Chew. E.** (2015). The Next Wave of Learning with Humanoid Robot: Learning Innovation Design starts with "Hello NAO". In T. Reiners, B.R. von Konsky, D. Gibson, V. Chang, L. Irving, & K. Clarke (Eds.), Globally connected, digitally enabled. Proceedings Ascilite 2015 in Perth (pp. CP:52- CP:56).

³³² <http://www.rightsgodigital.com/nao-robot-activist/>

³³³ <http://www.straitstimes.com/singapore/2-humanoid-robots-are-helping-out-in-pre-schools>

³³⁴ <https://techcrunch.com/2017/04/14/robot-tutor-musio-makes-its-retail-debut-in-japan/>

³³⁵ <https://www.sciencedaily.com/releases/2016/10/161024095238.htm>

companion, a robotic assistant to be installed in key hospitals are recommended. This can be piloted as follows:

- i. The UK government to facilitate and support the Universities-hospital-robotic industry collaborations. Universities and companies that have expertise in AI and robotics can partner with local hospitals to develop personalised robotic assistants.
 - ii. To initiate national pilot interdisciplinary projects supported by industries and corporates for Corporate Social Responsibilities. These can be carried out with selected or volunteered hospitals, universities and companies for the intervention.
 - iii. Project grant calls to be opened to support the AI in healthcare action research by key research and industrial funders.
 - iv. The establishment of a national Institution for AI and Robotics in Healthcare as a catalyst for excellence for national and international show cases.
3. The issues raised in point 1 above and the measures proposed in point 2 will need broader and inter-disciplinary stakeholder consultations and an in-depth results analysis of their educational, psychological and economic impact. The findings are exemplars, good case studies and lessons learnt for widening participations. This can get educators, students, parents, patients, medical teams, general publics and industry more ready for the AI intervention and be the first country in the global scene to embed robotic companion in a systematic phases and large scale. Only after these empirical evidence are available and facts have been established can we conclude how to move forward, especially as regards legislative and ethical measures for AI and robotics used in the UK.
4. In addition, a holistic framework of providing education and healthcare with AI based on the larger scale of experimental research are required. Without analysis on the human-robot interaction from interdisciplinary aspects, i.e. psychological, socio-economical, educational, medical intervention, the effectiveness of robots in classrooms and hospital is questionable. Future research and design should leverage more refined data analysis techniques such as learning and medical analytics to directly focus on interaction and conversation dynamics³³⁶.

C. The role of the Government

³³⁶ Wong, N. W. H., **Chew, E.** & Wong, J. S-M. (2016). A review of educational robotics and the need for real-world learning analytics. Paper presented at the 14th International Conference on Control, Automation, Robotics and Vision (ICARCV), Phuket, Thailand 13 November 2016. <http://ieeexplore.ieee.org/document/7838707/?reload=true>

1. As a skilled AI and robotic researcher, I disagree with the speculative view of AI and robots will substitute human as depicted in some Hollywood movies. Instead, it is the installation of a pair of angel's wings (or devil's) subject to the designer. I am well aware of the ethical and humanities debates and how far does it make sense to various industries when the AI and robots came to commercialisation and mass productions. The value of a robot with AI capabilities reflects the values of those who make it and use it. I would recommend that, therefore, the role of UK Government is the key influential and guardian to set the boundary of the 'personhood' of AI and robotics in real-life implications. Following the European Parliament Legal Affairs Committee³³⁷, the civil laws on the 'personhood' of robotics and AI from research and design, program and development to manufacturing and commercialisation ought to be defined and debated.
2. It is suggested that an enhanced educational programmes or curriculums need to be reflected from pre-school way up to higher education for developing graduates with the skills that will never be replaced by AI and robots. Since the industrial revolution, our students have been educated for being better skilled labours in the educational sausage factories. When these jobs are being taken by AI and robots, it is the time to reflect what knowledge and skillsets are belonged to human, truly human education. A national forum, in-depth study, or oral evidence can be carried for relevant experts to discuss all possible jobs to be taken by AI and those which aren't, and why in order to create public awareness and to influence educational policymakers' decision.
3. AI research and design, the robotics code of conduct and ethical implications for industrial, universities and general public are the foundation of the promising side of AI, the angel's wings. Thus, the government should call for regular consultations with stakeholders in all industries and expert groups consists of legal experts, robotics researchers, industrial leaders, economics, humanists, educationists, psychologists and relevant panels to iteratively establish general regulations for the country, and specific legal affairs to govern various industries and fields that may be varying. For instance, (1) whether to give robots 'personhood' status as argued by EU committee³³⁸, (2) to penalise unethical conduct in designing AI algorithms and robotic programs; or (3) to introduce robot tax to fund support for or retaining of workers put out of job by robots³³⁹.

³³⁷ <https://ec.europa.eu/digital-single-market/en/blog/future-robotics-and-artificial-intelligence-europe>

³³⁸ <https://www.theguardian.com/technology/2017/jan/12/give-robots-personhood-status-eu-committee-argues>

³³⁹ <http://www.reuters.com/article/us-europe-robots-lawmaking/european-parliament-calls-for-robot-law-rejects-robot-tax-idUSKBN15V2KM>

5 September 2017

Children's Commissioner for England – Written evidence (AIC0123)

Introduction

The Children's Commissioner has a statutory duty to promote and protect the rights of all children in England with special responsibility for the rights of children who are in or leaving care, living away from home or receiving social care services.

Independent of Government and Parliament, the Commissioner has unique powers of inspection and discovery to help bring about long-term change and improvements for vulnerable children. She is the 'eyes and ears' of children in the system and the country as a whole and takes on issues which no other organisation is better placed to tackle.

The impact that digital technology has on the lives of children and young people is one of the Children's Commissioner's priority policy areas. As this committee inquiry recognises, children born today will grow up in a culture that increasingly relies on AI-based technology. The Children's Commissioner believes that, like the world of social media before it, this technology offers great opportunities but children's needs should be carefully considered as it is developed.

Growing Up Digital

Growing Up Digital³⁴⁰, published earlier this year, was the Commissioner's call for a step-change in the way that children are prepared for digital life. While children and young people make up a third of internet users³⁴¹, policy-makers, parents and teachers are not adequately equipping children with the skills they need to negotiate their online lives. The Commissioner called for interventions from government in order to give children and teenagers digital resilience, information and power, and hence open up the internet as a place where they can be confident digital citizens. These interventions are designed to be adaptable to new technologies, such as AI.

Digital resilience

The Commissioner recommended that all children should be taught 'digital citizenship' from age 4-14 with a voluntary extension for older children who want to become digital leaders or champions with an emphasis on becoming responsible and aware digital citizens.

As children should be taught about the value of their data and online interactions

³⁴⁰ <https://www.childrenscommissioner.gov.uk/publication/growing-up-digital/>

³⁴¹ Livingstone, S., Carr, J. and Byrne, J. (2016). One in Three: Internet Governance and Children's Rights. Innocenti Discussion Paper No.2016-01, UNICEF Office of Research, Florence. Page 15

they should also be taught what AI is and how to identify where it is being used. There are already AI-based programmes that integrate into social networks to spot signs of mental health issues among young users – signs such as talk of suicide, self-harm or persistent negativity.³⁴² These programmes are able to identify these signs quickly and offer the user choices for support. While this is beneficial in providing early intervention advice, children should be taught how this technology works and where it is being used so that they are fully informed when they choose to disclose private information. It is also important for them to be able to recognise whether they are talking to a human or interacting with 'chatbot' technology.

Digital information

The Commissioner also called for more straightforward social media terms and conditions in language that children could readily understand. Similarly, children should have an understanding of both where AI technology is being used and how their data might be used, as well as how their rights interact with the rights of a company. As the Commissioner called for greater transparency from social media companies, the same should be applied for those developing AI programmes. As AI programmes build richer data profiles, children should understand that this data may be collected and in order to prevent these profiles being traded or exploited, it should be subject to greater protections than those offered in the existing Data Protection Act and forthcoming General Data Protection Regulation. In the UK, children's data does not currently have this level of protection within existing legislation.

Encryption that prevents AI programmes from sharing data with the social media platform it is built into would be one way to tackle this. Another would be to ensure that data gathered for the purpose of service delivery could not be transferred or sold to secondary, partner or other companies. Either way, greater transparency would ensure that children are appropriately informed at the point of their interaction.

Digital power

In Growing Up Digital, the Commissioner also called for a children's digital ombudsman to provide a route for children to seek reconciliation of issues relating to social media and to drive better accountability by service providers. When AI development becomes increasingly complex so does accountability for decision-making³⁴³. It is therefore essential that we provide children with

³⁴² See Woebot in the UK: <https://woebot.io/> or Bark in the USA: <https://www.bark.us/faq>

³⁴³ The Institute for Electrical and Electronic Engineers, 'How can we improve the accountability and verifiability in autonomous and intelligent systems?' https://standards.ieee.org/develop/indconn/ec/ead_law.pdf

protection, clarity of information and the power to question decisions made about them.

Where AI is being used in other areas of children's lives – such as in education to deliver more effective learning tools and to categorise children into the appropriate learning groups³⁴⁴ - it is important that these decisions are supported by human judgement. AI programmes can make inaccurate or unfair assessments of an individual where the training data available reflects existing human bias or is based on insufficient data coverage.³⁴⁵ Given this potential, children should always have the power to challenge decisions made about them by AI programmes.

A minimum standards framework

The Children's Commissioner recommends consideration of a 'minimum standards' framework for developers of AI in all its forms, to minimise any negative impacts of programmes on children and young people and help maximise the potential for AI. A similar approach has already been taken by the UK Centre for Child Internet Safety in their 'guidance for providers of social media and interactive services'³⁴⁶, however it was developed many years after the establishment of the major networks that dominate children's lives.

A new minimum standards framework for the development of AI would include:

- The collection and responsible use of children's data in relation to AI
- Inclusive youth participation in the design and development process
- How biased data can impact children and why it should be mitigated
- Children's ability to understand these systems/ processes, and to trace decisions back to ask questions about why and how a decision was made

Conclusion

The internet was not designed with children in mind yet it has significantly changed the way they interact and continues to be the source of new opportunities and risks for all children. While the DCMS's Internet Safety Strategy will likely lead towards a safer, more transparent internet and to more informed digital citizens, children would be in a better position if there had been a pre-emptive and systematic consideration of children's rights as the internet began to develop. As we enter a new wave of change driven by machine learning and AI, we have the knowledge and responsibility to take such a pre-emptive

³⁴⁴ Intelligence Unleashed: An argument for AI in Education, Pearson & UCL Knowledge Lab (2016): <https://www.pearson.com/corporate/about-pearson/innovation/smarter-digital-tools.html>

³⁴⁵ This is evident where machine learning has been used in predictive policing schemes: Human Rights Data Analysis Group Oakland study/ Lum and Isaac (2016), Royal Stats Society <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2016.00960.x/full>

approach. As outlined in the United Nations Convention on the Rights of the Child, children have specific rights as children and not just as an extension of adult rights. It is therefore crucial that when considering the implication of developments in AI, we deliberately consider children as an independent group and that their views and interests are taken seriously, particularly in relation to any AI-based technology that is directed at or used by children.

6 September 2017

CIFAR – Written evidence (AIC0136)

1. Artificial Intelligence (AI) technologies are increasingly performing human-level cognitive activities – from perception and recognition to decision-making and inference. AI technologies have the potential to play an important role in improving our quality of life; however, pausing to understand their effects, both unintended and intended, is crucial, as institutional and social adaption may be outpaced by the advent of breakthrough AI technologies. Developing and funding a comprehensive research agenda on the implications of AI in society is critical today.
2. As philosopher and CIFAR Advisor Daniel Dennett states in his new book, *From Bacteria to Bach and Back Again*, “I am not worried about humanity creating a race of super-intelligent agents destined to enslave us, but that does not mean I am not worried.”
3. We congratulate the Select Committee on Artificial Intelligence for this call for evidence and appreciate being invited to respond.

1. AI and CIFAR

- 2.
4. Since our inception in 1982, CIFAR has been committed to research excellence and impact in fields as diverse as artificial intelligence, inorganic photosynthesis, early brain development, and inclusive prosperity. We take a long-term view and support leading edge research with the potential for global impact. Our close to 400 fellows and advisors are drawn from 18 nations, including 20 from the UK.
5. We believe that research that focuses on critical issues, and leverages the best scientists and scholars, without being constrained by disciplinary boundaries or geographic borders, is the most effective way to create new knowledge for a better world.
6. Our research model is based on a combination of four mutually reinforcing characteristics: a focus on complex global challenges; a foundation in global, interdisciplinary networks; a flexible, open-minded approach; and sustained long-term commitment. Nowhere are the benefits of our unique approach for society more evident than in the field of AI.
7. CIFAR’s first research program was Artificial Intelligence, Robotics and Society. It was officially launched in 1984 and included many leading thinkers of the time, including Professor Geoffrey Hinton (University of Toronto, Google Brain).
8. In 2004, CIFAR launched a new AI program, Neural Computation and Adaptive Perception, under Professor Hinton’s leadership. The new

network was a key actor in keeping alive an approach to building intelligent machines that drew inspiration from neuroscience.

9. A few years later, the work of CIFAR Senior Fellows in this program developed into what is currently the dominant approach to AI, called “deep learning.”
10. This technique relies on artificial neural networks that learn from examples to make efficient and accurate predictions and decisions. Coupled with more powerful computers, large data sets and techniques to train deeper networks, the field of AI continues to grow dramatically as technology advances at a rapid rate.
11. Today, this CIFAR program, renamed Learning in Machines and Brains, is co-led by Professor Yoshua Bengio (Université de Montreal) and Professor Yann LeCun (New York University, Facebook). It continues to include many of today’s highly respected leaders in the AI field including two researchers based in the UK, Professor Nando de Freitas (University of Oxford) and Professor Christopher Williams (The University of Edinburgh).
12. In its most recent budget, the Government of Canada recognized CIFAR’s long-term commitment to AI research with an investment of CAD \$125 million (~GBP 77.5 million) in the Pan-Canadian AI Strategy.
13. Developed and implemented by CIFAR, this Canadian AI program has four mutually reinforcing objectives: to attract and retain outstanding AI researchers based at independent institutes in Canada’s main AI centres – Edmonton, Montreal and Toronto; to increase opportunities for graduate training in AI; to develop national and international activities for AI research and training; and, to create a program that catalyses an inclusive, global conversation on the implications for AI in society.
14. This year, to move forward on the fourth goal, we will be launching with others an initiative to support international working groups to examine the economic, ethical, policy and legal implications of advances in AI.
15. The implications of AI for society remain uncharted territory and require the urgent attention of researchers, policy-makers, industry and civil society if we are to ensure that societies benefit from AI. In partnership with leading organizations, we look forward to stimulating discussion and new insights about the opportunities and challenges of AI.

AI and Science

16. AI is having significant impact in other areas of science and scholarship. From materials science to biomedicine to energy to epidemiology and

economics, AI technologies are providing new ways to analyse data sets, identify correlations and predict potential outcomes.

17. Last spring, CIFAR hosted a workshop on Machine Learning for Energy Materials Science at the Massachusetts Institute of Technology in Cambridge, Massachusetts. The workshop brought together researchers from academia and industry based in Canada, the US and Europe to explore how AI could accelerate research into the discovery of new renewable energy materials, catalysis and storage.
18. Research into AI technologies is also profoundly changing human genomics research in areas such as bio-marker discovery, molecular diagnostics and genome-based therapies for complex disorders. In 2015, CIFAR Senior Fellow Professor Brendan Frey (University of Toronto) launched a start-up company, Deep Genomics, based on research showing that AI has the potential to diagnose gene mutations linked to human diseases such as cystic fibrosis and spinal muscular atrophy.

AI and Health

19. As the example noted above indicates, AI is on the precipice of transforming healthcare, most notably medical imaging and diagnostics. Recent advances in AI have the potential to reduce mortality rates due to medical errors and inaccurate diagnoses, while lowering the cost of diagnostic services. Recent research has shown that some AI technologies in the medical field have reduced error rates to levels akin to human-alone interventions.
20. As CIFAR Senior Fellow Professor Max Welling (University of Amsterdam) articulates, new AI technologies will change the role of practitioners as machines take on the more technical aspects of diagnosis and imaging. Moreover, wearable health-oriented technologies will become effective tools for individuals and healthcare practitioners alike.
21. The health care sector is already beginning to integrate AI technologies today and is planning for profound changes to the sector. One area that is often noted is the need to develop new approaches to training and recruitment for health care practitioners, with skill sets to complement the technological shifts in the industry.
22. Keeping patient data confidential must also be of utmost concern – for example, re-identifying anonymized data by matching datasets from different sources. Today, there are technical solutions being explored to resolve this issue.

23. In addition, visualizing large neural networks is immensely challenging, and understanding the requisite algorithms requires extensive technical training. This situation will improve by creating algorithms that help to explain computer predictions, and thus, improve the propensity for adopting AI technologies. There is also research underway to resolve this challenge.
24. Finally, the issue of correlation and causation must be considered. AI technologies will discover causal relations from observed data. It is hard to imagine that health care practitioners would not always need, in some way, to be responsible for analysing the data to understand whether there is a correlation or causation with certain illnesses.

AI and the Labour Market

25. One way to begin to understand the impact of AI on the labour market is to consider two ways that a given AI technology can be implemented with respect to workers – enabling and replacing technologies, as presented recently by CIFAR Senior Fellow Professor Daron Acemoglu (Massachusetts Institute of Technology) at a CIFAR keynote address in Ottawa, Canada.
26. Enabling technologies complement and increase the productivity (and wages) of certain types of skills (e.g., laptops for managers and workers specializing in problem-solving, scanners for cashiers). In contrast, replacing technologies conduct tasks previously performed by labour (e.g., assembly tasks, switchboard operation, mail sorting). This can further lead to displacing labour, reducing wages and polarizing employment.
27. Research by Professor Acemoglu suggests that replacing technologies do not always create long-term negative effects. As new machines replace labour in some tasks, new tasks in which labour has a comparative advantage will be created. In the long run, this can boost growth, generate wealth and encourage consumption.
28. However, disruption can be costly in the short-term due to job losses concentrated in certain sectors and their social implications. The attendant concern must be that job growth will increase economic inequality by preferentially impacting low-paying service jobs.
29. More research is required to understand how the potential of AI can be fully and equitably realized across society. One area that is often mentioned is education and (re-)training at all levels – from primary to tertiary – to equip the current and future labour force with the necessary skills (not just digital skills) to work with AI-assisted technologies. Funding social science research into these potential outcomes is critical.

AI, Ethics, Culture and the Law

30. It is hard to imagine a sector of society that will not be affected by AI. Sectors that are already experiencing major disruption include transportation (e.g. autonomous cars, ride-share services and delivery systems by drone); financial services (e.g. e-commerce and transaction support systems); and, communication (e.g. human communication using AI systems for spoken and dialogue agents, chat bots and machine translation services).
31. While we believe that these changes should be embraced by governments, it is essential that the ethical, cultural, regulatory and legal issues be thoroughly researched and understood by policy-makers, business and civil society.
32. In addition, governments should explore the need for both policy and regulatory frameworks to ensure that industry and others do not misuse AI. From privacy and consent to fairness and statistical bias to interpretability and transparency, the ethical and legal issues presented by AI technologies are complex and interconnected.
33. The social and economic issues are also broad and diverse – from understanding how humans and machine will interact to planning for the future of work and education to ensuring that a small number of companies do not monopolize the field due to their technological dominance.
34. Culturally, there are many questions that require interrogation by philosophers, artists and scholars such as the nature of consciousness and intelligence, the meaning of being human and the obligations, rights and responsibilities that arise from these considerations.
35. From our discussions with leading AI researchers, social scientists and policy-makers, it is clear that research in all of these areas is just beginning to get underway and is critically needed.

Conclusion

36. Based on our long and sustained support for research at the frontiers of AI, we believe that access and adoption of AI technologies should be encouraged by government, provided that it is coupled with a clear understanding of the relevant social, ethical, regulatory and legal implications.
37. Developing a research agenda and moving forward with new scholarship in all of these areas is critically and urgently needed. We also believe

governments should continue to fund fundamental research in AI, with an eye for the long-term betterment of humanity.

38. The impact of AI in society is a defining issue of our time and requires international attention, including in developing nations. As deeper understanding emerges, it must be supported with a clear policy action plan that ensures that AI technologies promote social good.

39. We commend the House of Lords Select Committee on Artificial Intelligence for initiating this public call for evidence. We look forward to working with partners in the United Kingdom as we move forward to address these issues and enhance understanding in the area of AI in Society.

From:

Dr. Alan Bernstein, President and Chief Executive Officer, on behalf of CIFAR

6 September 2017

Donald Clerk – Written evidence (AIC0022)

My father was born in Scotland, and because of this I was able to get a British passport. I am therefore, in some sense, a citizen of your country – although I have never lived in Britain. In any case, I welcome the opportunity to submit something to your committee because I believe that concerns about artificial intelligence go beyond borders. There is certainly some potential for AI to be beneficial for humanity, but, in my opinion, there is no other technology that have ever been invented that has greater potential to cause harm, either intentionally, or accidentally.

I am an electrical engineer by training; in practice, I've spent the last 20 or so years as a software test engineer working on telecommunications systems. I am not therefore an AI expert. But, by virtue of my training and work experience, I think I have perhaps a greater ability to understand some of the developments in AI than most ordinary people.

I suspect that what I have to say concerns two of your questions

3: How can the general public best be prepared for more widespread use of artificial intelligence?

5: Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Regarding question 3, I think the public (here in Canada at least) has no clue about artificial intelligence and my guess is that it is probably the same in Britain. In this regard, you need to understand that the public includes many technical professionals like me. Although I'm an engineer, and I've worked with software for many years, I had no clue about what is happening with artificial intelligence until about 2 years ago.

You should check what I'm about to say with some experts, but my intuition tells me that right now the future is being invented for people by the scientists, engineers, and entrepreneurs who create, develop, and deploy technology. Most of these people are trying to create things that are useful and safe – but of course it is their definition of what is useful and safe! And the people who create, develop, and deploy new technology have a vested interest in it. For the academics, it is their career; for the engineers it is their career; for the entrepreneurs, it is they who, more than anything else, stand to make a LOT OF MONEY from artificial intelligence. Innovation may well be an important engine of economic growth, but is also today, an engine of economic, social, and political disruption on a massive scale. And that disruption, while often beneficial economically, is also a cost on society. If for example, AI causes a lot of job losses to occur very rapidly, it will cause enormous suffering for people – even as

it creates new wealth and opportunities for others. Innovation creates opportunities but it also creates huge risks.

To counterbalance the viewpoint of the people who have a vested interest in developing technology, the general public needs to be able to participate in an intelligent and responsible way, in decisions about what humanity does with technology. We, meaning all of us, need to understand the impact technology has on our lives, so that the decisions about what we do with technology are made based upon what is good for people – not just what is good for business, or for the people who have a vested interest in developing and deploying new technologies.

In order for the public to participate in a meaningful way, it isn't necessary for everybody to suddenly start becoming programmers, or data scientists, or AI experts. It is necessary to become educated about AI (and technology generally) enough to be able to understand the effects this stuff could have on people in human terms. To understand this, it is almost better not to become too technically knowledgeable! The value of public participation will come precisely because the general public is not part of the science/technology/business community that is developing AI.

So, in answer to question 3, you need to find ways to educate the public about technology and the consequences of its use and misuse, either intentionally, or willfully. Again, check this with the experts, but I think you will find that any time we apply a technology there are always “side effects” - unintended consequences of using technology. Some of those consequences are good, but some are bad. You don't just get the good stuff when you apply technology – you also get some negative side effects. Negative side effects are to some degree intrinsic to technology. You can try to foresee, control, contain, and mitigate negative side effects, but you can't completely eliminate them.

The public (at least in Canada, although I suspect it isn't much different in any of the rich western democracies) has an almost naive and blind trust in technology and the people who create, develop, and deploy it. People need to have a more realistic attitude I think.

With respect to question 5, my recommendation is that you should look for ways to educate the public about technology – not just the how to part of it, the human impact part of it. This should be a required part of education for people in school prior to university.

In university, all people studying STEM (science, technology, engineering, math), and business students should be required to take a course in understanding the impact of technology and about responsible innovation. The course should be mandatory – if you don't pass the course, you don't get your degree. The course should be rigorous and of very high quality. It should not be what we used to call

Donald Clerk – Written evidence (AIC0022)

(in engineering school in Canada at least) a “bird course”.

I hope my submission will be useful to you, and that it will give you a few good ideas. Good luck in your efforts to prepare your society to grasp the opportunities that AI presents, while at the same time avoiding the many dangers widespread application of this technology will trigger - unless you manage AI wisely.

Donald Clerk P.Eng.
Canada

23 August 2017

CognitionX – Written evidence (AIC0170)

05 September 2017

Charlie Muirhead, Founder and CEO

Tabitha Goldstaub, Co-Founder and Head of Business and Community Development

4. What is the current state of artificial intelligence? How is it likely to develop over the next five, ten and 20 years?

- a. **News, media and entertainment:** AI adoption is high. AI has been taken up widely by both legacy players in this industry as well as newer social media entrants. There are three broad areas where AI is being used: to create content automatically across a wide range of media types, to analyse and manipulate content more effectively, and to personalise the media experience in many dimensions for the audience.
- b. **Smart cities, personal/public transportation and logistics:** AI adoption is high. A lot of the technology coming to market now is proven and well-funded. AI is set to become a defining force in these industries. Key applications for AI cover autonomous vehicles (including drones), transport planning, environmental sensors, public service management and automated deliveries, long haul and last-mile.
- c. **Manufacturing, design and industrial robotics:** AI adoption is high. While robots are increasingly relied upon within design and manufacturing, the advent of AI is enabling new capabilities. Accenture predicts AI will benefit the manufacturing sector more than all others, boosting gross value added by \$4tn (US dollars) by 2035. Some of the ways AI is changing manufacturing are through computer vision (allowing machines to perceive objects and environments), optimising scheduling, design and consumption for efficiency, predictive analytics for maintenance and repairs, AI-powered generative design, and improving human-robot interactions.
- d. **Legal and professional services:** AI adoption is high. Cognitive automation is already starting to impact firms and displace lower value tasks. Early adopters report regulation is a barrier to significant transformation. AI is being deployed across three key areas: document and contract creation support, document analysis, and robo-advisory services.

- e. **The online customer experience:** AI adoption is high. Significant investments continue to be made in the worlds of marketing, advertising, sales and communications. CB Insights assess this area as among the top five hottest markets for AI companies at the moment, based on how many equity deals are being made. Key applications are understanding customers and the wider market, automating and optimising outbound communications, automating customer service processes, enabling conversational interfaces, and empowering consumers with more information and choice. Building trust with consumers is set to be the key battleground in the future of this area.
- f. **Financial services:** AI adoption is moderate. The financial services industry is seeing a lot of AI activity and investment. AI-powered services are helping organisations become leaner and more evidence-based, as well as enabling highly personalised financial products, advice and customer support at scale. Common investments include chatbots, P2P lending, smart wallets, predictive analytics, automated processes and mobile customer support. Key applications for AI include surveillance and fraud detection, algorithmic trading and credit risk scoring.
- g. **Human resources and recruitment:** AI adoption is moderate. Much effort is going into improving the function of HR through AI. The start-up community is pushing the boundaries of what is possible (and perhaps acceptable) while some of the incumbents are starting to invest in AI for inclusion into widely adopted tools. The single biggest area of activity is recruitment, within which the most competitive markets are candidate sourcing and hiring process automation. Companies can also deploy AI to support their learning and development programmes. AI is also emerging as a tool for performance management - particularly automating objective setting and evaluation - but also being able to predict capacity and performance for individuals and teams.
- h. **Healthcare:** AI adoption is moderate. Major players in the market now include big tech companies like Google as well as major healthcare companies, investing significant resources to drive innovation. In the US, adoption is high: about 86% of US healthcare providers, life science companies, and technology vendors are currently using artificial intelligence technology (according to Tata Consultancy Services). Adoption is likely to be slowed by the complexity of the environment and privacy concerns. Key use cases for AI in healthcare include improving diagnoses through data analytics, wearables and medical imaging, new forms of

personalised treatment, drug discovery, and automating and optimising administrative processes in the healthcare sector.

- i. **Mental health:** AI adoption is moderate. There has been rapid expansion by numerous start-ups into this space, but these are still early days for AI in mental health. In particular, many companies are still building sound clinical backing for their products. Generally, the technology is not as sophisticated as in other sectors. There are three broad ways in which AI is being deployed within the mental health sphere: diagnosis, designing personalised treatments and monitoring the effectiveness of interventions. Reported benefits of popular solutions on the market include flexibility, convenience and preservation of anonymity.
- j. **Science and the environment:** AI adoption is moderate. While the latest AI techniques are familiar to researchers across these areas, the markets for products and services are still emergent. Key use cases include improving researcher productivity, optimising energy efficiency, monitoring wildlife and the environment, supporting agriculture, and improving weather predictions.
- k. **Cybersecurity:** AI adoption is moderate. Professionals are expecting, as AI technology develops, to see more automated and increasingly sophisticated social engineering attacks. One of the biggest risks is AI's ability to replicate human behaviour. However, AI can also be used as part of a package of defences - to detect anomalous activity, monitor risk and orchestrate remediation.
- l. **Insurance:** AI adoption is moderate. Some incumbent firms are well underway with programmes to evaluate the potential of AI. Already, one-fifth of insurance carriers have invested in machine learning techniques with 42% planning to invest in the near future, according to Celent. Start-ups' have an excellent opportunity to disrupt as new commercial realities take shape. AI opens opportunities to gain extra insight from previously dark or unstructured data - for example from previously unseen trends, social media, browsing behaviour, or data from reinsurers. Additionally, the use of AI can help increase the accuracy of the underwriting process, improve fraud detection, claims management, marketing and the customer experience. One of the most lively issues surrounding AI, which the insurance industry will grapple with more than most, is bias and the potential for AI to exaggerate systematic discrimination.
- m. **Politics and government:** AI adoption is low. Even the richest governments have been slow to make the most of available AI technologies. While governments can be unique sources of the

quality data needed for machine learning, and the opportunities for public benefit are radical, implementation is rarely straightforward. Complex data consolidation, cleansing and privacy expectations have to be met. Significant potential exists for governments in many similar ways to the private sector. Key applications include the automation of routine tasks, easing frontline customer service functions by using chatbots, and making better, more informed decisions at all levels - from policy modelling to case work. AI is also having an impact on politics itself. Prominent use cases are the psychographic microtargeting of online advertising and political research to understand public sentiment and make electoral predictions.

- n. **Social good and humanitarian crisis response:** AI adoption is low. The latest AI technologies have huge potential in this sector but adoption has so far been limited. So far it is only possible to point to a handful of examples of charities and not-for-profits deploying AI to support their work. This sector faces particularly acute challenges when it comes to being able to capitalise on developments in data science, both in terms of the datasets available and the skills to develop the solutions. It is worth noting that many of the most useful AI applications for charities and not-for-profits are not unique to their sector. We can, however, identify some use cases which have particular importance within the third sector: understanding the underlying causes of health and social issues, monitoring and evaluating impact, fundraising, making predictions about issues before they arise, and collecting information about crises to coordinate responses in the most efficient way.
- o. **Education:** AI adoption is low. It is relatively early days for the adoption of AI in education, but it is attracting strong interest both in the academic and in the business community. There are some exciting start-ups emerging. Key applications include personalised learning, new forms of content delivery, online assessment and automation of educational administration.

• What are the ethical implications of the development and use of artificial intelligence?

- It is worth noting there is broad public support for AI. According to a survey by ARM Northstar, 61% of people are confident in AI's ability to add value to society. However, there are a number of concerns about the effect AI will have on jobs, social cohesion, and other aspects of human society.

- **Jobs:** In the same survey, 57% of respondents were concerned about AI becoming more intelligent than humans and replacing their jobs. Consensus is building that job displacement will be a likely consequence of this new generation of AI technology - set to replace repetitive, less creative tasks, and working its way up the 'value chain' as AI becomes more and more sophisticated. On the whole, AI is set to replace tasks, not jobs. It will be common to see employees working much more productively using AI, or fewer people performing the same work. We are optimistic about the creation of new, higher quality jobs but we have real concerns about 'pinch points' in the workforce where automatable tasks constitute most, if not all, of the job - one prominent example being professional drivers. We would encourage massive public investment in supporting these workers to retrain and be redeployed, preferably before they are unemployed.
- **Data protection:** In the same survey, 85% of respondents stated they were concerned about their privacy and the security of their personal information. AI often requires collecting and processing vast volumes of data about people's lives and personalities. Of course this raises issues of consent on an individual level: 'am I willing to hand over this data in exchange for more relevant services?', as well as issues of power at an aggregate level: 'what does data monopolisation and abuse look like and how do we stop it?' Generally GDPR is seen as a helpful balancing force in these tensions, although it is widely noted the regulation was never designed to address issues arising regarding AI, and ambiguities remain.
- **Trust:** Trust in how an organisation uses and protects personal data will develop into a more decisive market dynamic. Companies will thrive or fail based on whether they get this right. Often our community highlights the importance of the 'cool versus creepy' continuum - when is the ability for AI to anticipate individuals' needs cool and when is the fact that an algorithm seems to know you better than you know yourself creepy? For AI to drive social progress companies and public bodies will need to get this balance right. A complementing public education programme about the capabilities and potential of AI also seems essential.
- **Anthropomorphisation:** A common question we come across is whether AI should pretend to be human. Just because a company can AI capable of passing the Turing Test doesn't mean it should. It is a popular opinion in our community that it is wrong to overly humanise these machines, despite the marketers' analysis that it helps sell them better - see Amazon's Alexa, or Pepper robots.

Companies deploying chatbots for customer services have already found customers are forgiving of mistakes when they know it's artificial, but feel tricked when systems pretend to be human. We also regularly hear concerns about gender stereotyping based on the function of the AI. One common example is the choice of female personas for AI secretaries.

- **Decision-making:** Our community has come across thousands of variations of the 'trolley problem' and the list seems endless. In the absence of regulation, the ethical frameworks of the system's developers will win out. To ensure ethics are front and centre of AI projects organisations must include resource dedicated to considering ethical issues. This could include ethics experts assigned to AI teams as well as an ethics office that can work at the strategic level to ensure the best outcomes for the organisation and society.
- **Bias and discrimination:** One of the biggest concerns when it comes to AI is working with socially unacceptable bias. Any prejudices and inequalities we have as a society can end up coded into our systems. One of the reliable ways we know can mitigate this risk is to have more diverse development teams in terms of specialisms, identities and experience. Particularly regarding gender, this is a huge challenge; few young women take up technology subjects and careers; just 16% of the graduates in computer studies are women and the figure is 14% for engineering and technology. Nearly all of the 200-plus senior women in tech who responded to a recent survey had experienced sexist interactions.

13 In what situations is a relative lack of transparency in artificial-intelligence systems (so-called "black boxing") acceptable?

We find it helpful to define the question as to what level an AI is explainable. This then demands a clearer definition of what an explanation constitutes. It is likely to need to vary on a case-by-case basis. In fact, this is one of the gaps when it comes to GDPR. Most of the time it's perfectly possible to explain the general working of an AI system, and the input data used, but it might be tough (if not impossible) to audit its step-by-step working.

It is also necessary to approach this problem by looking at outcomes. What do the systems' decisions look like in the aggregate? Consider the example of a racist judge: one can audit their reported decision making processes all day but what really matters is the outcomes - whether they're locking up people equally.

There is a promising area in using AI systems to monitor the decisions of other ones - particularly for where velocity or volume would be beyond human capability - for instance in the case of monitoring autonomous cars.

- What role should the Government take in the development and use of artificial intelligence in the United Kingdom?

For the UK Economy

There is an urgent need to increase support of domestic AI powered business to ensure we can home grow world leading products. SEIS and research grants go some way but these need to increase if UK companies are to stand alone and resist the acquisition offers from the US and China. We suggest the following additional activities:

- **Government spend on UK AI:** Use government procurements to pump prime the UK supply base. Mandate that government departments spend a prescribed percentage on solutions from UK qualifying AI start-ups. This way the government can support the UK to develop leading global technologies.
- **Bring clarity and connection:** Use AI to understand this fast-moving sector. Find ways to make it simpler for digital businesses looking to deploy AI to find companies, software products, hardware and data sets that they might need.
- **Support businesses to navigate adoption:** Partner to build a free advisory chatbot to provide guidance to companies deploying AI so they can find out about the rules and regulations that apply to them.
- **Diversity in AI:** Encourage companies to reflect the balance in society, hire a diverse workforce and ensure they are building solutions for everyone.

For The Population:

- **Lifelong learning:** Promote skills and retraining so we can see redeployment, not unemployment.
- **Public education:** Explain the realities of AI and promote mass understanding not mass hysteria.
- **Ethics discussions:** Create a space for the ethics discussion to be had by the everyday person and include them in the proposed Data Use and Ethics Commission.

6 September 2017

Cognitive Finance Group – Written evidence (AIC0010)

- 1 Cognitive Finance Group is a consultancy specialised in applied artificial intelligence in financial services, and the views expressed here reflect our company's views.
- 2 We are the trusted adviser to Boards and senior management on scoping, selecting and implementing artificial intelligence for business growth and increasing competitive advantage.
- 3 We have a dedicated team of strategists in financial services as well as data scientists and machine learning experts.
- 4 We bring practical knowledge of applying artificial intelligence in financial services gained from client work, work which is informed by our over 45 years experience in blue-chip global organisations and our specialist knowledge about artificial intelligence.
- 5 Michael Aikenhead is a Director at Cognitive Finance Group. He completed bachelors degrees in computer science and law followed by a PhD examining the application of AI for automating legal reasoning. Michael headed the European knowledge engineering team for an AI company which was acquired by Oracle and has spent over 10 years working in banking.

Financial services in the UK economy

- 6 Recent Parliamentary material indicates that Financial Services
Contributed £124.2 billion in gross value added (GVA) to the UK economy,
Contributed 7.2% of the UK's total GVA
There are over one million jobs in the financial and insurance sector
Provide (3.1% of all UK jobs)
The UK had a surplus of over £60 billion on trade in the financial

What is AI?

- 7 Tesler's theorem asserts that "AI is whatever hasn't been done yet."
- 8 We adopt a pragmatic definition that AI is the creation of computer programmes that 'Do something that would require human intelligence to perform'.

Disruption to the UK workforce from AI

- 9 Influential study suggests 47% of total US employment is vulnerable to automation from AI technologies (Frey and Osborne 2013 and Bowles 2014 in a European context).

- 10 However much lower estimates also suggested e.g. 9% in OECD countries (Arntz, Gregory and Zierahn 2016).
- 11 Disagreements over size of effects reflect automation criteria applied and particularly whether examination is at the occupation or task level (Chui, Manyika and Miremadi (2015) estimate that 45% of work activities could be automated using already demonstrated technology).
- 12 Similar uncertainty on employment replacement or displacement effects (Petropolous 2017).

Survey of workforce economists concerning the impacts of AI on employment by 2025 indicates 48% envision displacement of significant numbers of both blue- and white-collar workers
However 52% expect that technology will not displace more jobs than it creates by 2025

Notably it is easier to see where automation might do away with the need for human labour than where technology might create new jobs – but it has always been like that - imagine trying to tell someone a century ago that cybersecurity specialists would be important or that e-sports competitors may be the next big sports stars.

- 13 Regardless of exact quantum of impact, it is clear that there will be workplace change and potentially large scale change.
- 14 Notable that by lowering the skills required to perform a task and work, the pool of available people who can perform the task is also expanded, as are the locations where the work can be performed.

Financial services business model pressures

- 15 Competitive pressures from
 - Existing financial services competitors
 - Fintechs
 - International technology companies
 - Consumer expectation for on-demand personalised service
- 16 Existing UK financial services organisations will increasingly feel necessity to improve the efficiency and personalisation of service and product delivery.

Job creation in the AI economy

- 17 Studies into impacts of AI in the workplace suggest certain types of activity will be less subject to automation, those involving:
 - Perception and Manipulation; Creative Intelligence; Social Intelligence
- 18 Jobs vary in the degree they can be automated:
 - Managing others; Applying expertise; Stakeholder interactions;
 - Unpredictable physical work; Data collection; Data processing;
 - Predictable physical work

- 19 The automation of existing work through application of AI will create new jobs and reshape existing jobs. For example:
 - Algorithms and associated tools will need to be built, trained and maintained
 - Business specialists will need to put new tools to best use and to manage business models and processes that use the tools
 - Management and governance will be needed of tools to interpret and explain results and ensure their fairness and acceptability
- 20 As the workforce reshapes the UK has opportunity to capture a share of the jobs that are created
- 21 There is potential to further develop expertise in AI, and the application and management of AI
- 22 It is an open question though whether jobs created as business models reshape, will be actually be created in the UK
- 23 For example much activity is by large non-UK global technology companies
- 24 While the UK is currently seen as a global leader in AI this needs to be maintained and leveraged to ensure financial services business model and workforce changes are positively navigated
- 25 We do not want to see a situation in which financial models and tools are built outside the UK and merely consumed here with the lightest possible footprint
- 26 The financial services industry needs access to the skilled staff with AI knowledge and experience in order to thrive
 - Unfortunately we are already hearing about shortages of skilled data science staff

Governmental responses

- 27 Establishing cities or countries as hubs for AI development requires joining the global competition to attract AI talent and investment.
- 28 It is an International marketplace so the UK must be internationally focused.
- 29 UK has become a leader in financial services technology, but will need to continue to invent and innovates the tools that will serve other markets.
- 30 UK needs to maintain the ecosystem to allow this to technology development to happen

Encourage and Support Innovation

- Invest in university research
- Support industry R&D
- Encourage academic and industry collaboration

Venture Capital Financing

- Encourage the continued availability of VC funding
- Support expansion of available VC funding

Develop Talent

Weight formal education with STEM skills used in AI innovation

Encourage the best students working on AI to come to and remain in the UK

Encourage the best professional talent to come to the UK and work on the technical and business innovations that will occur in financial services business models

Work with financial services companies to encourage and support their workforce reskilling (e.g. Salesforce is examining ability of companies to turn themselves into universities so they can educate in-company as people get displaced – transforming into a digital workforce)

Lead with Government as a model user

Support innovation through adoption of tools

Strong and Stable Governance

Innovation minded regulatory framework

Predictable Regulatory Framework

Access to Markets

Support export of financial services technology and expertise

9 August 2017

Competition and Markets Authority – Written evidence (AIC0245)

Does the CMA envisage the principles underpinning Open Banking ever being applied to other non-financial data?

Much of our work on Open Banking following our Retail Banking Market Investigation has been about putting in place the necessary infrastructure and removing barriers so that the principle of Open Banking can be seen through in practice. We saw the potential of an idea that already existed and enabled / expedited it to achieve better outcomes for consumers in this particular market.

Some studies on open banking have had a wider vision of an 'API ecosystem' that would encompass not just payment services but savings and investment products and other financial services such as insurance. We are aware that app developers are already working on the use of bank transaction data outside financial services, for example to provide advice on choice of energy and telephony/broadband supplier.

The principles underlying Open Banking are similar to the new portability principle in [GDPR](#) –and there is a lot of potential in the portability principle to help get data working for consumers. Whilst each market is separate, in general we do see that the principle of data portability could be very powerful for consumers in a range of markets. For example, in our energy market investigation, we similarly made recommendations to help harness the power of customer data to improve outcomes for energy users.

The Committee might find it useful to look into [developments in Australia](#) with regard to giving people access to their own data in energy and telecoms etc.

What implications are there for the use of AI due to the introduction of Open Banking?

AI may be viewed as a new (and cheaper) way of delivering tailored advice derived from transaction data. It will use algorithms based on the scripts currently used by sales people which help them to know their customer. There is potential for AI / algorithms / data to be used in tandem with Open Banking to deliver new services which could help consumers. The incentives for banks to invest in AI is huge as it transforms the economics of 1 to 1 advice. It allows personalised advice and responses to questions at virtually no marginal cost and with near-perfect control of the content imparted.

General Information on Open Banking

Open APIs are central to our package of remedies from our retail banking market investigation.

The Open Banking remedy has the potential not just to reduce or remove the frictions that customers encounter on their existing 'journey' of searching for, selecting and potentially switching providers, but to change nature of the customer journey itself by facilitating the emergence on a large scale of new service providers with different business models offering innovative solutions to consumers and SMEs.

The development and implementation of an open API standard for banking – our core foundation remedy – will permit authorised intermediaries to access information about bank services, prices and service quality and customer usage. This will enable new services to be delivered that are tailored to customers' specific needs. The types of new and improved services that will result from this remedy include applications which:

- Allow banking customers, through a single application, to manage accounts held with several providers.
- Allow customers to authorise the movement of funds between current and deposit accounts to help avoid overdraft charges or to benefit from higher interest payments.
- Let customers make simple, safe and reliable price and service quality comparisons tailored to their own usage patterns.
- Monitor a current account and forecast a customer's cash flow, helping to avoid overdraft charges.
- Use a small business's transaction history to allow a potential lender other than their bank to reliably assess the business's creditworthiness and offer better lending deals than they would without this information.

Some third-party services already exist which demonstrate the potential new options that open APIs would make available to banking customers. These include, for example, services which monitor transactions and balances in current accounts, forecast the account holder's cash flows and provide a line of credit (or a link to alternative lenders) whereby money is automatically paid into the account if it is necessary to do so to avoid overdraft charges and withdrawn subsequently when the account is back in credit.

However, to use these and similar services it has generally been necessary for customers to disclose to the service provider their internet banking log-in credentials which may affect, or be perceived to affect, the guarantees against fraud that banks provide. This inhibits take-up and we believe that such services will gain greater market acceptance when our remedy, which removes the need

for customers to disclose these highly sensitive details to a third party, is adopted.

APIs are the key to the digital services that are used on computers and smartphones. They enable users to share information, for example on location or preferences. They are the technological drivers behind digital applications like Facebook, Google Maps and Uber. In banking, APIs can be used to share, in a secure environment, information such as the location of bank branches, prices and terms of banking products. APIs may also be used, with the customer's informed consent, to share securely their transaction history to enable access to tailored current account comparisons and other services.

We are requiring the largest retail banks in both GB and NI to develop and adopt an API banking standard so as to share information to a specified timetable and we are requiring it to be an open standard so as to enable it to be widely accessible. This will enable intermediaries to access information about bank services, prices and service quality. Customers who are satisfied about privacy and security safeguards, and are willing to give consent, will be able to share their own transaction data with trusted intermediaries, which can then offer advice tailored to the individual customer. This will make it easier for customers to identify the best products for their needs.

For information on how open banking is being implemented see:

<https://www.openbanking.org.uk/about-us/>

20 December 2017

Contact Centre Systems Ltd. – Written evidence (AIC0032)

HOUSE OF LORDS

SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

Submission by Mark Tindal, owner of Contact Centre Systems Ltd. and an independent consultant specialising in the business application of Artificial Intelligence in the Customer Services industry. I was published in 2012, accurately predicting the current rise in an AI based service industry, particularly in Contact Centres. I have 25 years experience in all areas of Contact Centre and Collaboration technology.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- 1.1 University degrees teaching Machine Learning started to become more popular around 20 years ago, students leaving academia between 2000 and 2005 were keen to move that technology from the classroom to commercial use. One such use was the **replacement of human agents in the Contact Centre** where humans often do little more than read from scripts.
- 1.2 Early AI around 2010 using Semantically Indexed databases or other methods of mimicking human cognition performed well but needed “tuning” by Computational Linguists, skilled and trained to understand the human conversation. Adoption by businesses was poor as a result of the effort required to bring a “Virtual Assistant” up to the level of even the most basic human agent.
- 1.3 Fuelled by the launch of Apple’s Siri and it’s relatively successful adoption by the general public, many organisations took this as a sign that there could and would be a successful commercial model for AI. This led to the second and third generation of AI, Cloud based with masses of compute and storage power.
- 1.4 Consumer interaction with intelligent machines will continue to gain rapid pace over the next 5 to 10 years and beyond as global businesses like Amazon advance their use of speech as a more natural input medium.
Significant advances in Speech Recognition should be considered as

important as AI in your discussions. The easier the conversation is with a machine, the quicker the general public will adopt it. “Natural Language Processing” is on the same curve of rapid growth and the general public have taken very well to the second stage of active listening in Amazon’s Alexa, the first being Siri and Cortana from Apple and Microsoft respectively.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

2.1 Yes, AI should be taken very seriously when one considers the ease with which a machine can respond to a customer question on a website. Right now, that answer has to be programmatically added to the machine but we are now at the stage where technology is learning how to answer questions based on their human counterpart in the Contact Centre. Once enough business knowledge has been consumed by intelligent machines there will be very little need for humans with lower levels of education.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

3.1 Referring again only to the Customer Services industry, in my opinion there will be a **reduction in the number of humans required to interact with customers of around 40% by 2020 and that will rise to 70% by 2025.** Humans answer around 8.15bn calls to UK Contact Centres of which there are 7,500 employing just under 1 million people. HMRC and the Department of Work and Pensions are the largest with 13,000 and 28,000 respectively. That’s 400,000 then 700,000 people who will need to reskill or employ their knowledge in other parts of the business.

3.2 Those human agents with specific business knowledge **can be employed as internal consultants to the technical community** as AI based systems are installed and as they evolve over time. It’s unlikely that an entire Contact Centre would be retained for this purpose, however.

3.3 Contact Centre agents who have at least rudimentary knowledge of their sector and not selected by their organisation may not fully understand the value associated with that business knowledge. **These individuals would benefit from a Government backed training scheme to help them apply their sector knowledge and secure a consultative role in a technology business.** These individuals would also be ideal ambassadors for the adoption of AI technology. Somewhat ironic I know. I’ve been considering launching something along these lines for some time and would welcome central government support.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

5.1 In my opinion there is little to be gained by Central Government education of the public in Artificial Intelligence. With the greatest respect, it's my experience that public sector does not move quickly enough to react to the pace of change in technology.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

6.1 All businesses with Contact Centres are driven to reduce the costs associated with connecting a customer to a human, so significant investment in "Virtual Assistants" is now very common. Repetitive and simple tasks are easily replicated by machines either on a web browser or via a speech recognition system making Contact Centres an ideal breeding ground for the advancement of Artificial Intelligence.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

8.1 Intelligent Machines are likely to replace the unskilled and semi-skilled in our offices first; those performing repetitive and simple tasks will no longer be required and may struggle to find alternative employment. This would place a tremendous burden on our Social Care system.

8.2 Under current EU legislation HR departments have the right to make business decisions that result in the reduction of their workforce. While there is an obligation to try and find suitable employment in other parts of their business, there is very little responsibility on an employer to find an "at risk" employee alternative employment elsewhere. The impact on Social Care budgets could be reduced if businesses were given tax incentives based on their outgoing staff working practises.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

- 9.1 The basis of all Artificial Intelligence is a combination of standards from academia, primarily Statistical Modelling and Machine Learning to create a cognitive model of sorts. **No one organisation or group of organisations has the rights to this theory.** It's true that organisations with vast quantities of working capital are investing heavily in their own versions of these models to solve problems they believe capture the essence of AI but academics are constantly refining the base models.
- 9.2 The result is a constantly and rapidly changing underlying methodology that is unlikely to grant enough competitive advantage to any business, regardless of budget. However, it is likely that one or more of the world's biggest business will assume dominant positions and attempt to artificially create a standard based on the changing standards.
- 9.3 **The closest and most simple comparison would be that of the universal serial bus (USB).** Whilst there is a standard "universal" USB pin configuration, voltage and current, set out in the same way; there are now six different ways it can be presented, eight if you include Apple's version.
- 9.4 The presentation layer of Artificial Intelligence will be ultimately important and it's likely that before 2020, developers will be designing and building applications to one or at most two competing standards.
- 9.5 As it stands currently, developers are given the perception of a lot of choice when it comes to creating basic consumer bots, for example wit.ai, api.ai, Amazon Lex provide access to their service creation environments for free, but the functionality is very limited and in my opinion only intended to wet the technical communities appetite.
- 9.6 Regardless of the emergence of one, two or even more premium AI service creation environments, developers will always be access to newer 'garage builds' as the technology matures. **The issue that could be address by government is one that was left un-checked in the Speech Recognition industry for 30 years.** The hostile acquisition, consumption and break-up of smaller competing technology businesses resulting in a monopoly.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

10.1 The vast scale of the practical applications of Artificial Intelligence would make it virtually impossible to regulate, however Central Government should take all steps necessary to ensure that the Competition Commission work as a matter of urgency on a clear definition of their powers. Government cannot allow dominance of Artificial Intelligence in any sector by any one or any group of businesses. **Perhaps government could consider enrolling the major players in AI to a strategic alliance with clearly define terms of reference** that include the protection of human interest and the advancement of technology.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

29 August 2017

Cooley (UK) LLP – Written evidence (AIC0217)

Introduction

1. This submission is made by Cooley (UK) LLP, the London office of the international firm Cooley LLP. Its author is Mark Deem, a partner in the firm, who advises and is regularly consulted on matters concerning issues of liability in such areas as data, cybersecurity and artificial intelligence.
2. Cooley is a firm of over 900 lawyers across 12 offices in United States, China and Europe, who solve legal issues for entrepreneurs, investors, financial institutions and established companies. Clients partner with Cooley on transformative deals, complex IP and regulatory matters, and high-stakes litigation, often where innovation meets the law. Cooley supports artificial intelligence companies (or those using artificial intelligence technology) at all stages of development, from company formation and capital financings, through privacy and information security concerns and intellectual property and licensing support.
3. In May 2017, Cooley teamed up with the Institution of Engineering and Technology (IET), Peter Warren and Cyber Security Research Limited to host an international artificial intelligence conference in London, which brought together key thought leaders in artificial intelligence to explore the ways in which artificial intelligence will have an impact on the way we conduct our lives, our work and leisure. A further conference is scheduled for March 2018, developing on these themes.
4. Cooley is also the legal partner and a co-sponsor with Google of the Research and Applied AI Summit, a platform for entrepreneurs and researchers who accelerate the science and applications of artificial intelligence technology.
5. This submission necessarily focuses on those areas of the Call for Evidence of the Select Committee of July 2017, which require a response from a legal perspective.

Pace of Technological Change

6. Artificial Intelligence lacks a uniform definition and accordingly may be used to refer to a wide sphere of technological activity and may embrace a wide variety of meanings depending upon the context in which the term is used.
7. For the purposes of this submission, the term “artificial intelligence” or “AI” is used to describe a product or technology, which generates an

action in response to its perception of the environment in which it operates. This definition deliberately incorporates two core aspects of AI – the detection and collation of ever more data points from a wide variety of sensors at ever-increasing speeds (so-called 'big data' sets) and their interpretation, often by automated or algorithmic means.

8. AI is presently being deployed – to varying degrees of technical complexity – in a wide number of industries. The pace of change has been significant over the past five years and this is expected to accelerate over the coming five and ten year periods (and beyond).
9. From a legal perspective, the pace and success of development will be intrinsically linked to whether an appropriate and suitably robust legal framework can be developed, which will support growth, encourage investment in this technology and allow risks to be fully understood and properly provided for, whether technically or through more traditional risk management mechanisms, such as insurance.
10. Factors which will inhibit the pace of change include the failure of society to engage with unresolved questions of accountability as a matter of law; the imposition of an inappropriate or oppressive regulatory regime; and issues of cybersecurity which have the ability to impact the quality and integrity of data captured for interpretation. These points are expanded below.

Transparency in development of Artificial Intelligence

11. There are presently four key areas, where the existing legal framework in this jurisdiction would benefit from detailed consideration, in the context of artificial intelligence technologies, to support their development and to harness their maximum potential:
 - a) Legal liability – the basis upon which legal liability can be established in respect of an artificial intelligence technology;
 - b) Issues of causation and accountability – the basis for determining which party is to be considered liable (or is prepared to accept liability) for artificial intelligence, which does not perform as expected;
 - c) Use of AIs in seeking to perform or discharge existing legal obligations; and
 - d) Legal status – the extent to which a legal status should be afforded to an AI.
12. We identify (in outline only for present purposes) each of these issues, before identifying some areas for consideration.

Legal Liability

13. As artificial intelligence technology develops, it will challenge the underlying basis of legal obligations according to present concepts of private law (whether contractual or tortious).
14. As a matter of legal theory, contractual obligations have at their heart the concept of a bargain – a promise supported by some valuable consideration. These obligations require the promisor to deliver on the promise – nothing more – and, failure to do so, results in the promisor being liable to put the promisee in the position it would have been in, had the contract been performed. It is essentially a form of strict-liability between the contracting parties, which assesses performance against a promise at the time that promise was made. Absent further agreement, the nature of the promise will not change.
15. By contrast, tortious obligations have that their core, the imposition (or assumption) of a duty between two parties considered in law to be “proximate”. They require a party to discharge the relevant duty, without causing the proximate other any harm. Failure to do so, results in the wrongdoer being required to provide redress for the harm caused. It is a form of fault-based liability, which assesses loss when the harm occurs.
16. If we now consider the proper performance of an AI product or technology, which may involve the incorporation of a black-box algorithm (or even a data set whose true integrity and credentials are unknown), it is clear that the transparency deficit in the algorithm creates legal uncertainties based upon present concepts of legal liability.
 - a) From a contractual point of view, the presence of artificial intelligence makes it difficult to define the promise and therefore the bargain with any certainty. Indeed, perhaps the only aspect which has any certainty is what goes into the ‘black box’. However, in circumstances where the essence of any contract is the value provided by the algorithm, any promise which is confined to what goes into the black box is very limited in nature. Further, any upside which goes above and beyond such a limited promise, as a matter of pure contract theory, would be for the promisor not the promisee.
 - b) From a tortious point of view, the proliferation of open source (black-box) algorithms – absent some very real evidence of “proximity” – is likely to lead either to the inadvertent imposition of a duty on a party or significant difficulties being encountered in establishing a duty and therefore liability.

17. Neither of these positions is ideal and so the status quo is ill-equipped to support development: to the extent that any party seeks to use an open source (black-box) algorithm, it is unclear how any valuable contractual promise based upon outcomes can genuinely be made and the party might unwittingly be assuming a liability to persons unknown.
18. Consideration therefore needs to be given to two particular areas: first, whether those engaging in this area should be encouraged and incentivised to define the nature and extent of their involvement and to offer, where appropriate, greater transparency as to how the AI concerned operates; secondly, whether a more appropriate legal basis for liability might be closer to concepts of fiduciary law.
19. Unlike contractual obligations, fiduciary duties are not necessarily chosen but imposed on an individual and may change in nature as the relationship between the fiduciary and beneficiary evolves. Any additional upside achieved by the fiduciary enures to the beneficiary, rather than the fiduciary. (It is noted that, in this regard and without further agreement, any benefit over and above the promise made under a contractual obligation will be for the promisor rather than the promisee).
20. Whilst it is certainly not proposed that those involved in AI should be treated as pure fiduciaries, encouraging the development of artificial intelligence by those who have a view of and assume some responsibility towards the ultimate user could form a basis for a framework.

Issues of causation and accountability

21. In terms of causation, the use of artificial intelligence technologies in products which then go on to cause a loss, raises the question of how we will readily be able to determine who, as a matter of strict law, is to be held properly accountable – especially in circumstances where one or more parties might have contributed to the loss.
22. Implicit within the definition of artificial intelligence set out above is the combination of big data sets and algorithms. This leads to a proliferation of potential parties accountable for the loss, beyond the designer and manufacturer of the core product or software, to include the designer of the algorithm, the coder of the algorithm, the implementer/integrator of the algorithm, the owner of the data set that has been interpreted, the creator of the original data point, and so on.
23. In turn, this presents a legal challenge as to how one can assess the relative merits of the various positions of “stakeholders” in the loss and opens up the prospect of extensive litigation based on claims and counterclaims of the various parties seeking to avoid, or minimise, liability.

24. Encouraging parties to define the precise parameters of their own liabilities at an early stage – defining (in so far as they are able) the extent to which liability is accepted or assumed – will enable a sensible ring-fencing of liability, where risk can be managed and, where appropriate, insured.

Use of AI to discharge existing legal obligations

25. A number of obligations presently permeate commercial agreements and involve parties warranting a certain state of affairs exists (or will exist) or requiring performance of relevant obligations in a certain manner, whether using best endeavours, reasonable endeavours or exercising reasonable skill and care.
26. These obligations will almost certainly have been assumed without consideration of how they can properly be discharged and indeed whether use of artificial intelligence could achieve such discharge. Can it be said that using technology, which incorporates a black box algorithm, would meet the standard required to discharge a best endeavours obligation? Or is a sufficient response to an obligation requiring reasonable skill and care to be deployed?
27. As matters stand, courts can expect to receive substantial evidence seeking to establish the point. Consideration should therefore be given as to whether the imposition of certain quality processes as to the integrity of data sets and on those seeking to design, code or implement algorithms so that parties are able to take suitable steps to facilitate the discharge of legal obligations.

Legal Status

28. In early 2017, the Committee on Legal Affairs of the European Parliament made recommendations to the European Commission on the Civil Law rules concerning robots. One particular recommendation, which captured the imagination of the wider public, was whether a specific legal status should be considered for the most sophisticated autonomous robots.
29. Whilst this particular recommendation is arguably beyond the scope and mandate of this Select Committee as presently constituted, it nevertheless raises significant legal questions as to how we are going to deal with the legal validity of transactions carried out by robots on behalf of their owners – for example, in whose name does the transaction take effect; or who is the ‘inventor’ of an AI-produced product, over which IP is being asserted.
30. Whilst it is conceded that a discussion of these issues is welcome, it is important that any decision to create a legal personality recognises the

primacy of humans and that such legal personality is a deemed legal fiction, for convenience.

31. Analogies can be drawn in this area to the manner in which Roman law developed to allow and grant limited rights and transfer of limited powers to slaves in a manner in which their masters retained overall responsibility.

Role of Government

32. In circumstances where the present legal environment is challenged by a number of issues arising from the development of artificial intelligence, taking no action is not a realistic option.
33. As a general proposition, however, the executive and legislature should exercise caution in seeking to provide for how technology should develop. Legislators, with the greatest respect, are not technologists, nor should they seek to be.
34. Any regulation should be expressed in terms of a framework, rather than seeking to be too prescriptive, and should have the ability to flex in a nimble manner (through case law and secondary legislation) to meet the ever-evolving environment and applications of artificial intelligence.
35. Widespread and early adoption of any framework should be encouraged by engaging with the widest constituency of those involved in artificial intelligence (whether users, investors or those impacted by the technology) at an early stage.
36. Finally, regulation of artificial intelligence cannot be considered in isolation, but in conjunction with both the integrity and security of data and the wider internet of things.

12 September 2017

Dr Steven Cranfield, Chrissie Lightfoot, Michael Butterworth, Ms Joanna Goodman and Dr Paresh Kathrani – Written evidence (AIC0104)

Dr Steven Cranfield, Chrissie Lightfoot, Michael Butterworth, Ms Joanna Goodman and Dr Paresh Kathrani – Written evidence (AIC0104)

[Submission to be found under Ms Joanna Goodman](#)

Will Crosthwait – Written evidence (AIC0094)

About the author

Will Crosthwait is CEO of Fintech AI company Kensai - Artificial Intelligence using news and social opinion to gain financial insights. He also advises on AI, is taking part in a roundtable on AI between the Royal Society and the London Mayors Office and recently delivered a talk on AI at the BBC. Evidence is supplied in an individual capacity.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1 Artificial Intelligence is currently in a highly exciting and fast paced environment due to the emergence of IBM Watson API's and Google DeepMind and TensorFlow open-source library. These allow big business, small SME's and entrepreneurs to access the power of IBM and Google AI platforms for relatively small cost, which will lead to a proliferation of AI companies. This access for small teams will see companies applying solutions to problems in almost all industries. Over the next 5 years, AI will enter the larger industry sectors like healthcare, finance, education and media before moving into social and entertainment sectors over the next 10 years. As the technology scales and solidifies up to 20 years, we can expect to see transformational technologies which change society and potentially, even what it means to be human.

1.2 The availability of API's will accelerate the development of AI as large tech companies like IBM, Google, Microsoft and Amazon compete to capture the infrastructure market of processing the World's data. Governments will have to decide whether to try to regulate and control the AI or nurture a nascent market that could ultimately discover solutions to many of society's ills.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

2.1 The current level of excitement surrounding AI is warranted but is also muddled and confused. Modern AI, machine learning and deep learning, will bring about the greatest technical and societal change since the beginning of the industrial revolution, greater than everything the digital age has bought so far.

2.2 Classical AI, namely algorithms, have been around since the 1950's and many companies using it are jumping on the Modern AI bandwagon by claiming

to be an 'AI startup'. While technically true, the conflation of Modern AI and Classical AI means the true value of Modern AI is being obscured. Though as access to Modern AI through API's becomes more widespread, many companies will adopt Modern AI and the true value of AI will be revealed.

2.3 As private companies, and hopefully government, realise the efficiency savings and problem solving ability of AI, there will be a massive impact on society. Jobs where people assess images, offer scripted support or analyse data will most likely become automated. As the service industry dominates almost will be automated or augmented to be delivered using AI. This either frees up the labour force to concentrate on more creative problems and less menial tasks or means retraining a workforce to participate in more manual labour.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

3.1 Artificial Intelligence will soon be capable of delivering a highly personalised experience for its users, accessed through voice control and powered by data. This means jobs regarding the gathering or analysis of this data will grow massively. But for people employed in jobs where they are currently offering that service, they will need to be retrained.

3.2 When assessing industries that will be impacted it helps to try to understand types of data and how broad this is when comprehending AI. Data can mean a conversation, medical symptoms, an educational curriculum, stock markets, the rule of law or public opinion across social media and the news. All can be analysed, learnt and the services around them automated to a lesser or larger extent by AI.

3.3 There are two schools of thought around what this means. There's an optimistic movement that sees monotonous tasks being eliminated and freeing up the workforce to do more creative work like managing human relationships. The more pessimistic viewpoint is that work will be replaced at a rate of change that means people cannot be retrained fast enough, one that ultimately sees the introduction of a universal basic income.

3.4 The creation of an AI economy and society is likely to have both pros and cons for the public but as with the industrial and internet revolutions, the jobs lost to technology are likely to be offset by the opportunities created.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

4.1 People doing manual labour or jobs where human interactions are required are likely to only feel positive effects of AI. For a proportion of those in industries

that are effected by AI, it will be felt as enormously positive, with a large range of new tools that help take the monotony out of their jobs and allow them to increase the number of customers, patients or students they can serve. For those in low skill service jobs, it is likely that retraining will be needed to provide new skills.

4.2 The potential solution to any labour offset by AI is artificial intelligence itself, as it could be used in an educational setting on mobile phones or computers to retrain people or give them new skills.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

5.1 Organisations like the BBC are already making efforts to improve the public's understanding of AI with initiatives like the Blue Room's season on AI and automation. As private companies educate the public on the unique selling points of their AI products, people will naturally begin to engage with and understand AI.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

6.1 The key sectors that will instantly benefit from AI will be healthcare and education. AI is already being used to accurately diagnose health conditions in apps like Babylon Health and large organisations like the NHS will be able to gain huge cost saving from the data efficiencies and analysis that AI will bring. AI will also be used to deliver education, whether in the form of a curriculum or a highly tailored learning experience that is likely to be overseen rather than taught by a teacher.

6.2 The industries that benefit fastest are going to be those that engage and interact with SME's that are already leveraging value from AI rather than waiting for larger corporate solutions.

7. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

7.1 An important distinction to be made when assessing the safeguarding and availability of data is that of anonymised data and data that contains private information. Some of the largest sources of data in the UK are collated or owned by the government and are increasingly being digitised to be made available. Huge breakthroughs can be made using government owned anonymised data and this fact could be the big differentiator between the UK and other nations, like the USA, where the data is in the hands of private organisations (especially in areas like healthcare).

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

8.1 The internet age has already forced society to look at data and the use of data with considerable control measures already in place combined with enforcement from the Information Commissioners Office and from May 2018, the General Data Protection Regulation.

8.2 The largest unanswered ethical questions are around areas like crime and justice, where decisions made using data can have huge impacts upon people's lives. Northpointe's tool, called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), is using AI to carry out risk assessment of recidivism for the US justice system. A report by Propublica³⁴⁷ looked at more than 10,000 criminal defendants in Broward County, Florida and found that black defendants were often predicted to be at a higher risk of recidivism than they actually were while white defendants were often predicted to be less risky than was shown. Sceptics of the system believe this could be due to institutional racism in the existing data.

8.3 When dealing with such sensitive areas, it's critical that the data used for learning is clean data, that is potentially weighted (although this introduces ethical questions of its own).

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

9.1 There are already systems being established to give the right to an explanation of the algorithm used from initiatives like GDPR and its associated rights related to automated decision making and profiling. In areas like credit scoring it will be possible to have a human look at your claim if it is refused by

³⁴⁷ <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>

AI. While it will prove almost impossible to see the actual steps taken by AI in an individual case, unless a solution comes forward, it isn't unreasonable to know general data points that have been used to make the decision. Although there will have to be a careful balance between openness for users and protection of IP for the technology companies.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

10.1 The Government should take a supportive role in the development and fermentation of British artificial intelligence startups. Ideally, they will learn from the lessons of the startup boom in the early 2010s and instead of giving vast sums of money to accounting firms like Grant Thornton to consult to the startups as with 'Growth Accelerator' or to pay marketing firms to create a 'Tech City', the funds should be given directly to SME's in the artificial intelligence fields to maximise benefit to the nation. This can be done under existing State De Minimis aid rules. Regulation or lack of government funding is likely to stifle growth, especially when considering competition in the form of China. It would be great to see the government create a specific grant fund for AI companies to maximise on the opportunity while incentivising individuals to learn about AI.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

- 11.1 The European Union's policy approach toward AI is likely to be similar to that of its approach to internet businesses with GDPR and the Horizon 2020 funding programme being the result. Navigating EU funding opportunities has always been time consuming and favours large companies that have large teams but potentially lack agile innovation. Suggested policy around 'personage' from the recent EU Parliament Legal Affairs Committee on Robotics is likely to be able to answer many questions around taxation of AI but drawing the line as to what level of AI is considered a person will prove to be a challenge. The biggest lessons are likely to be learnt from China, who hope to beat the US to become the first AI superpower by 2030. The citizen-centric privacy legislation coming from the EU around data, which is arguably the fuel that feeds the AI machine, is unlikely to be matched by China where large amounts of public data will increasingly become available to Chinese companies. China wants to generate 400 billion yuan (\$59 billion) of AI related output per year by 2025 as outlined in their "Next Generation Artificial Intelligence Development Plan" (新一代人工智能发

展规划) which also calls for establishing initial frameworks for laws, regulations, ethics, and policy.

4 September 2017

Darktrace – Written evidence (AIC0243)

Written submission to the House of Lords AI Committee from Dave Palmer, Director of Technology, on behalf of Darktrace

1. Use of terms in this response:
 - a. **Artificial Intelligence (AI)** will mean the focussed application of machine learning approaches and other advanced mathematics for the solving of specific narrow problems, rather than a general intelligence (based on a human or otherwise) that can solve arbitrary problems.

Question

(1) What does artificial intelligence mean for cyber security today, and how is this likely to change over the next 10 years? Does artificial intelligence have implications for conventional cyber security today?

2. The fundamental value of AI enabling computers to deal with increasing complexity and subtlety is in principle a very significant benefit to cyber security defenders, who in general are not directly responding to the actions of external attackers, but who are fighting to overcome their business's own complexity, diversity and scale.
3. The challenge of defenders meaningfully understanding and monitoring a business, where every person and device behaves in a unique way and are often numbering in the tens or hundreds of thousands and spread across the country or globe, is realistically already beyond the ability of typically sized security teams.
4. Asking these teams to recognise the repeated patterns of previously known historical attacks is reasonable using conventional computing and software techniques (e.g. firewalls, anti-virus). But asking these teams to identify the strange and the out-of-character actions that might be the hallmarks of a novel attack or a disaffected employee, is not achievable personally (due to scale and inability to guess everything that might go wrong), and not achievable using standard software programming approaches (due to overall complexity, and subtlety of daily behaviours). This is why attacks can emerge within a business, escalate for months or years, and become a crisis without anyone knowing about it until the criminals decide to reveal the crime.
5. AI has an enormous role to play in exponentially improving existing protective approaches like anti-virus and firewall approaches, and also in enabling fundamentally new approaches like business immune systems that can learn the normal behaviour of everyone and everything, and respond to subtle changes indicative of emerging threats that are already within the business but that can be handled before they become a crisis.

6. Realistically, there are no other scientific developments on the horizon that can facilitate cyber security adapting to the ever-increasing scale, diversity and complexity of digital businesses in the foreseeable future. Without AI, cyber security doesn't have a hope of coping with inevitable digital growth.

Question

(1a) Does AI facilitate new kinds of cyber-attacks, and if so, what are they? Are these potentially more dangerous or threatening?

7. AI techniques will open up new opportunities for criminals to operate at greater scale and to pursue new models of criminality. Here are some examples:
8. First, imagine a piece of malicious software on your laptop that can read your calendar, emails, messages etc. Now imagine that it has AI that can understand all of that material and can train itself on how you differently communicate with different people. It could then contextually contact your co-workers and customers replicating your individual communication style with each of them to spread itself.
9. Maybe you have a diary appointment with someone and it sends them a map reminding them where to go, and hidden in that map is a copy of malicious software. Perhaps you are editing a document back and forth with another colleague, the software can reply whilst making a tiny edit, and again include the malicious software. Will your colleagues open those emails? Absolutely. Because they will sound like they are from you and be contextually relevant. Whether you have a formal relationship, informal, discuss football or the Great British Bake Off, all of this can be learnt and replicated. Such an attack is likely to explode across supply chains. Want to go after a hard target like an individual in a bank or a specific individual in public life? This may be the best way.
10. Second, imagine attacking all the connected smart TVs and video conferencing systems installed in meeting rooms across a target organisation, say a Law Firm. Typically, these devices are substantially inferior to the security of a modern laptop. Then activate the microphones and stream the audio of all meetings held, to an AI driven translation and transcription service (already available from Google and Amazon). Given transcripts of all meetings an additional simple AI model could automatically alert the criminal to topics of interest (perhaps unannounced Mergers & Acquisitions, or the details of preparations for a particular trial) and suddenly the criminal has easily scalable approaches for ambient surveillance of a company without having to actually listen to any meetings themselves. Ambient surveillance has previously been the stuff of spies not because it accesses uninteresting content, but because it doesn't scale very well, and AI completely changes

that whilst we as an economy are busily engaged in the sprinkling of our environments with cameras and mics.

11. Third, data has become a proxy for the beliefs of an organisation. Further in the future than the previous two examples (which are achievable now), imagine deliberately altering data so that an Oil&Gas firm's executives bid for drilling and mining rights in a wrong and not profitable location. Or a series of random bank account balances are subtly and consistently tweaked in the bank's digital backups before being changed in operational systems resulting in an inexplicable set of books and major loss of consumer confidence. Such attacks are more elaborate and would rely on smart software that was able to manipulate data in a manner that is believable at first glance but that becomes disruptive at scale. It's not unreasonable to believe this is achievable in the next 10 years.

Question

(2) To what extent can AI help to strengthen cybersecurity? Where are such approaches used in cybersecurity, and how might this change in the future?

12. Covered in Question 1 above.

Question

(3) Will only state-sponsored hackers have the means to deploy AI in cyber-attacks? Or is there a risk that AI-enabled cyber-attacks will be democratised in the near future? Does this make a difference when attempting to defend against AI-enabled cyber-attacks?

13. AI is absolutely democratised to anyone with a laptop and an internet connection. A motivated 'hobbyist' software programmer could almost certainly start from no understanding of AI, to deliver the types of attack described in the first and second examples above within 6-12 months. A more focussed criminal would be able to achieve this sooner and it is somewhat surprising this hasn't happened already.

Question

(3a) Are particular applications of AI, for example in healthcare or autonomous vehicles, more vulnerable to cyber-attacks than other areas, or is the threat quite evenly distributed across sectors?

14. In general, digital criminals and some nation states are mostly focussed on generating profits. Whilst it is easy to imagine that disruption to society could be horrifying (e.g. disrupting vehicles or interrupting critical healthcare), it is far more likely that the criminals would seek to generate maximum profits

whilst remaining sufficiently low profile that they don't become a primary target for law enforcement. It is perhaps interesting to be able to extort vehicle owners for a ransom before they can continue their journey but this is more a reflection of how motivated a person would be to pay a ransom in different situations rather than inherent weaknesses in the technology of different sectors.

15.No comment on the risks from espionage and hostile states / state affiliated groups. Insight on this is better sought from the UK Security & Intelligence Agencies.

Question

(4) Do AI researchers need to be more aware of how their research might be misused, and consider how this might be mitigated before publishing?

(4a) Are there situations where researchers should not publish or release AI research or applications with a high risk of misuse?

16.The types of AI research like image/person recognition and natural language understanding have tremendous momentum globally and for the most part offer significant benefit to society including offering exciting opportunities for healthcare. However, exactly the same technologies could be reused for criminality. We have discussed the ambient surveillance risks of speech-to-text processing above. Far more viscerally, real time video processing and image classification could allow a machine gun attached to a simple robot to target certain groups e.g. supporters of a particular football team, people of a particular race, or gender, without a terrorist being present nor digitally connected to a weapon they have left behind. Unfortunately, in these cases the "hard part" is the AI theory and code which are typically freely published and improved by thousands of individuals around the world. Incorporating these AI techniques into a program designed to cause harm is comparatively easy.

17.We cannot imagine a mechanism for controlling the publication of research that could mitigate some of these problems, as in the above example, the AI needed to power a self-aiming gun (or drone) would essentially be the same as that used to identify family members in a photo.

Question

(4b) Should the Government consider mechanisms, voluntary or mandatory, to restrict access in exceptional cases, in a similar way to the Defence Advisory Notice system for the media, for example?

18.We are not experts in this sort of control, but suggest it is likely more effective to rely on (existing?) rules that prevent the use of digital and physical weapons rather than to attempt to qualify AI algorithms or tooling as particularly good/harmful.

Question

(5) How much of an issue are recent developments in the field of adversarial AI for the wider deployment of AI systems?

(5a) Should more attention be paid to adversarial AI attacks when developing new AI applications?

19. The meaning of “adversarial” regarding AI is in flux around the world, but it is assumed that in this context it is intended to mean the deliberate training of weaknesses into an AI system so that it misclassifies or fails to identify certain objects or properties.

20. Developments in the thinking about adversarial approaches to AI are essentially identical to security researchers looking for flaws in computer software. The more awareness, understanding and activity we have on this subject, the better overall the outcomes of future AI development will become.

21. The historic boom in computer software, websites, and apps was not matched by a simultaneous boom in understanding of security vulnerabilities and has resulted in our current cyber security situation worldwide. If we can avoid repeating that mistake in the AI space, we will all be better off.

Question

(5b) Should mandatory regimes of stress-tested or penetration testing, prior to the release of systems or products, be required?

22. As with cyber security, a continuous approach to assess risk and stability over the lifetime of a product is far more preferable than a one-off hurdle that gets jumped over at the beginning. However, it is almost certain that the overwhelming majority of AI applications will be based on frameworks similar in nature to today’s Keras, Tensorflow, Caffe etc, developed by open source communities and tech companies like Google, Amazon, Facebook, Microsoft, Apple. This means that a focus of AI safety into a relatively small number of frameworks (and related educators) is likely to achieve a significant impact on the safety of AI decision making globally.

Question

(6) How prepared is the UK for the impact of artificial intelligence on cyber security?

(6a) Are the UK’s national institutions sufficiently protected?

23. AI is simply the next phase of the cyber security “arms race”. It has already been true for years that the businesses and organisations that most effectively resist cyber security crises are those that can quickly adapt to

modern security practices: including continuously modernising their technology generally, and more recently, incorporating AI in their cyber defence.

24. This is always uneven across the economy. Those organisations that struggle to modernise have been, and will continue to be, disproportionately impacted by cyber-attack events. Note that they are not attacked more, instead they suffer more when hit by attacks. The publically announced disruption to the NHS by the ransom attack nicknamed Wannacry, and the continuous loss of personal payment details by the low-profit-margin retail sector are *expected* examples of organisations who are not in a position to modernise their technology rapidly and so attacks escalate into crises more often.

25. There are no signs that this trend will be different as we move into AI-enabled attacks and defences.

Question

(6b) Is the National Centre for Cyber Security doing enough?

(6c) Should the National Cyber Security Strategy take explicit account of the threats and opportunities of AI for cyber security?

26. We do not have visibility of specific actions of the NCSC on this topic. However, the current state of cyber security in the NHS, and the recent public disclosures of parliamentary usage of their organisational IT and rife password-sharing suggests that the NCSC has considerable work to do on the basics with critical UK organisations before they get onto advanced themes.

Question

(7) Once the GDPR and the Data Protection Bill have come into force, will the law be able to adequately prosecute those who use misuse AI for criminal purposes?

27. No, these bills appear to be focussed on the punishment of organisational victims of cyberattacks if they are seen to have behaved in a negligent way in collecting, holding and defending personal information. We are not aware of any initiatives to improve cross-international-border law enforcement collaboration in a way that would facilitate the routine investigation and prosecution of criminal development or use of cyber-attacks (whether including AI or not).

Question

(8) How can personal data be effectively secured against misuse, especially given the potential conflict between secure and open data? Does the increasing availability of AI have implications for securing this data?

28.Covered in Question 1/1a above.

Question

(8a) How does artificial intelligence affect the security of anonymised datasets? Is there a level of anonymisation that is 'secure enough' to protect personal data against misuse?

(8b) Are provisions in the Data Protection Bill sufficient to ensure that cyber-security researchers are able to test AI applications and data anonymisation protocols, without fear of legal prosecution?

29.No comment. Not our area of expertise.

Question

(8c) Is there a role for blockchain, and distributed-ledger technology more generally, in protecting personal data from AI-enabled cyberattacks?

30.Blockchain and distributed ledgers may enable greater trust in the **authenticity** of stored data, but they will not offer any enhancements in the privacy/security of that data. Indeed, the requirement of many participants with access to the full content of data means that it is likely that data privacy overall will be reduced. It is the nature of cyber-attacks that even if the blockchain software itself was implemented perfectly, the rest of the required operating systems, software, hardware, and the human maintainers offer weaknesses and they will be more numerous than usual due to the distributed nature of the ledger around multiple participants.

Question

(9) How can we maintain the security of AI systems, particularly those of a safety-critical nature, both now and in the long term?

31.Security as a continuous process throughout the life of a product/service that is supported by consistent modernisation is vital. Questions 1/5/6 above have described how security needs to develop over the coming years and how continuous research into the abuse of AI decision making (via adversarial techniques) remains vital and needs to be mainstreamed for all AI developers.

Question

(10) Who should be responsible for securing and patching these systems, and how long should this responsibility be expected to last?

32.The owners of systems/products must be responsible for them and their security risks throughout their entire lifetime regardless of AI content or not.

Question

(11) What is the one recommendation you would like to see this Committee make with its final report to the Government?

33.AI offers the opportunities to supercharge both cyber defence effectiveness and, sadly, the speed, scale and automation of cyber-attacks. It is crucial that the Government maintains or increases its efforts to raise awareness of increasing and emerging cyber security risks to companies (currently via the National Cyber Security Strategy).

18 December 2017

Data & Society Research Institute – Written evidence (AIC0221)

Response to UK House of Lords Call for Evidence

Select Committee on Artificial Intelligence

M.C. Elish, Intelligence & Autonomy Initiative, Data & Society Research Institute

Overview

In this document, we respond to the Select Committee on Artificial Intelligence's call for evidence, and address a series of questions within the categories of (1) defining AI, (2) the pace of technological change, (3) impact on society, and (4) ethics. Our responses are based on qualitative research studies conducted by researchers and fellows at Data & Society,³⁴⁸ a non-profit research institute based in New York, and the work of our colleagues in academia and other social science research institutes. We thank the House of Lords for the opportunity to provide evidence.

Our definition of AI

1. What is AI?

We define "artificial intelligence" as a characteristic or set of capabilities exhibited by a computer that resembles intelligent behavior. What counts as intelligence is not any specific set of traits, but rather defined in relation to existing beliefs, attitudes, and technological capabilities. Computer scientists Russell and Norvig have written that the history of AI can be seen as variously emphasizing four possible goals for "intelligence": "systems that think like humans, systems that act like humans, systems that think rationally, systems that act rationally." Either than rely on a specific definition of AI, or even intelligence, we take the nebulous definition of AI has a central characteristic. Social *perceptions* of AI are as consequential as the official definitions and must be taken into account as AI technologies are examined and regulated.

³⁴⁸ Data & Society is a research institute in New York City that is focused on social, cultural, and ethical issues arising from data-centric technological development. To provide frameworks that can help address emergent tensions, D&S is committed to identifying issues at the intersection of technology and society, providing research that can ground public debates, and building a network of researchers and practitioners that can offer insight and direction. To advance public understanding of the issues, D&S brings together diverse constituencies, hosts events, does directed research, creates policy frameworks, and builds demonstration projects that grapple with the challenges and opportunities of a data-saturated world.

The pace of technological change

2. Is the current level of excitement which surrounds artificial intelligence warranted?

To see the effects of AI we do not need to look into the far future. The problems that need our attention are those that have to do with the mundane, and often invisible, ways in which AI technologies are proliferating in every sector of society. It is imperative to begin considering the implications of AI technologies on societies around the world because AI has the potential to change nearly every aspect of life, including the natural environment. Though it is important to consider a range of scenarios and timelines, it is also important to keep in mind that when public or policy-making attention is turned to the unlikely extremes of technology, like “superintelligence” or the “singularity,” this comes at a cost; attention focused on such issues distracts us from AI technologies currently being used, and the complex problems with which we are already confronted.³⁴⁹

Another important aspect of AI is its fuzzy and amorphous definition, as we noted above. In interviews conducted with those working in the deployment of AI systems, we found this to be a consistent theme.³⁵⁰ From product engineers to venture capital investors, AI was acknowledged as “the scariest but also the most universal” term.³⁵¹ The concept of AI was often leveraged to drum up excitement or stand in for a range of automated technologies. And it was defined differently by people in different contexts and situations. This may be useful in marketing and pitching new ideas, but it can have unintended consequences for how the public understands what a technology is – and what its limitations are. We have observed that non-expert understandings of AI are often shaped by marketing rhetorics suggesting capabilities that are not yet technically possible.³⁵² This aspect of AI is not “a bug” but rather, “a feature;” the fluctuating understandings and perceptions of AI cannot be resolved, but rather efforts should be made to account for the consequences of potential misunderstandings of AI.

Impact on society

³⁴⁹ Crawford, K., Whittaker, M., Elish, M.C., Barocas, S., Plasek, A., Ferryman, K. (2016). The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term. Report prepared for the AI Now public symposium, hosted by the White House and New York University’s Information Law Institute. Retrieved from https://artificialintelligencenow.com/media/documents/AINowSummaryReport_3.pdf

³⁵⁰ Elish, M.C. & Hwang, T. (2016). An AI Pattern Language. New York: Data & Society Research Institute.

³⁵¹ Ibid: 12.

³⁵² Elish, M.C. and Boyd, d. Situating Methods in the Magic of Big Data and AI. Communication Monographs. October 2017. Doi: 10.1080/03637751.2017.1375130

The implications of AI will be far reaching, and are impossible to comprehensively predict. A first principle of assessing AI's impact on society should be an acknowledgement that there will be no clear finish line or final resolution. The work of thinking through impacts on society must be a continuous investigation, and while solutions to problems should be sought they must never be understood as providing definitive answers.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

We believe that the most productive ways for the general public to be prepared for widespread use of AI will be to understand the limitations – alongside the possibilities – AI technologies, and also to begin grappling with the nuanced ways in which AI is effecting everyday life. Here, we highlight two ways in which everyday lives are likely to be impacted by AI in ways that could subtly but substantially benefit some more than others.

- Work practices and management

AI will change existing power dynamics between employers and those who work for them. As systems become “smarter,” management tasks are being allocated to algorithms. From shift scheduling to routing directions to hiring platforms, roles previously played by mid-level managers are being performed by AI systems. These technologies often promise rational and objective efficiency, often under the name of “disruptive” innovation.

One result of these disruptive technologies is that they destabilize or even undermine existing worker protections.³⁵³ AI technologies often do not fit neatly into existing regulatory models and their impacts have the potential to go unnoticed until too late. Practices of “smart” and “just-in-time” scheduling may benefit corporations' quarterly earnings, but ultimately result in work schedules that are unbearable for workers.³⁵⁴ For example, after the consequences of “smart” scheduling became widely known in the United States through the efforts of investigative journalists, researchers, and activists, many states within the U.S. have created new regulations to protect workers against unfair scheduling practices.³⁵⁵

³⁵³ Kneese, T. and Rosenblat, A. and Boyd, d., Understanding Fair Labor Practices in a Networked Age (October 8, 2014). Open Society Foundations' Future of Work Commissioned Research Papers 2014. <http://dx.doi.org/10.2139/ssrn.2536619>

³⁵⁴ Carrie Gleason and Susan Lambert, “Uncertainty by the Hour,” Position paper. Open Society Foundation Future of Work Project (2014). <http://static.opensocietyfoundations.org/misc/future-of-work/just-in-time-workforce-technologies-and-low-wage-workers.pdf>

³⁵⁵ Lam, B. Will States Take Up the Mantle of Worker Protection? January 17, 2017. *The Atlantic*. <https://www.theatlantic.com/business/archive/2017/01/worker-protection-schneiderman/513182/>

In addition, when companies such as Uber describe themselves as technology companies, not managers of employees, the conditions and implications of management shift. Such companies view their role as providing a marketplace that facilitates connections but that does not “employ” workers. If workers are not defined as employees, a company such as Uber is no longer responsible to workers as a traditional employer would be. However, in many ways companies like Uber present only a “mirage of the marketplace,”³⁵⁶ and ethnographic cases studies demonstrate that ride-hailing apps, such as Uber, structure and enforce particular modes of work just as a manager would.³⁵⁷ Ultimately, in this configuration, workers in the United States end up assuming the risks of employment without guaranteed benefits (such as decreased tax burdens, healthcare, and other workplace protections) or potential modes of redress.

If AI technologies are allowed to bypass existing norms and regulations because existing frameworks cannot adequately take them into account, this is likely to benefit corporations at the expense of individual workers, especially those in already vulnerable and precarious positions.³⁵⁸

- Access to opportunities and protections

One of the most potentially beneficial aspects of AI is also one of its most perilous: the capability to finely tune systems to individuals. AI systems are likely to be utilized in areas with significant decision making power over people’s lives, for instance in employee hiring,³⁵⁹ judicial sentencing,³⁶⁰ insurance

³⁵⁶ Hwang, T. and Elish, M.C. The Mirage of the Marketplace. Slate.com. July 27, 2015. http://www.slate.com/articles/technology/future_tense/2015/07/uber_s_algorithm_and_the_mirage_of_the_marketplace.html

³⁵⁷ Rosenblat, A. and Stark, L., Algorithmic Labor and Information Asymmetries: A Case Study of Uber’s Drivers (July 30, 2016). International Journal Of Communication, 10, 27. <http://dx.doi.org/10.2139/ssrn.2686227>; Calo, Ryan and Rosenblat, Alex, The Taking Economy: Uber, Information, and Power (March 9, 2017). Columbia Law Review, Vol. 117, 2017; University of Washington School of Law Research Paper No. 2017-08 <http://dx.doi.org/10.2139/ssrn.2929643>.

³⁵⁸ Citron, D.C., “Technological Due Process,” *Washington University Law Review*, vol. 85 (2007): 1249-1313.

³⁵⁹ Rosenblat, A. and Kneese, T. and Boyd, d., Networked Employment Discrimination (October 08, 2014). Open Society Foundations' Future of Work Commissioned Research Papers 2014. <http://dx.doi.org/10.2139/ssrn.2543507>

³⁶⁰ Rosenblat, A., Wikelius, K., Boyd, d. Gangadharan, S.P. & Yu, C. (2014, October). Data & Civil Rights: Criminal Justice Primer. Data & Civil Rights Conference. doi: 10.2139/ssrn.2542262; Barocas, S., Rosenblat, A., Boyd, d., & Gangadharan, S.P. (2014, October). Data & Civil Rights: Technology Primer. Data & Civil Rights Conference. doi: 10.2139/ssrn.2536579

coverage³⁶¹ or access to credit³⁶². While AI systems have the potential to benefit individuals in these circumstances, without careful attention to existing structural inequalities and biases, these systems have the very real potential to reinforce and perpetuate inequality.

Take the case of bias in hiring. AI systems present a potential way of addressing this issue. However, it is incorrect to assume that an AI system will automatically be fairer and unbiased.³⁶³ This is in part because these systems are built by humans – with their own biases and culturally-specific assumptions – and also because these systems rely on existing datasets to make predictions. Datasets are likely to reflect past biases or trends that were produced by previously biased or unfair practices.³⁶⁴ Often it is more subtle than simply accounting for something like gender bias. For instance, one “talent management” company found a correlation between job retention and the distance an applicant lived from her workplace. However, the firm also realized that including this correlation in making hiring assessments might unfairly advantage people who were able to live near work (which might be in a high-rent neighborhood, or a neighborhood far from a given ethnic community), disparately impacting disadvantaged socio-economic groups. Considering the potential for discrimination, the company decided not to include the metric in their system. This example underscores that correcting bias and protecting workers does not happen without keen attention, and sometimes a metric that seemed valuable at the outset might need to be discarded altogether.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

- Diversity is vital

As Crawford has pointed out in the *New York Times*, artificial intelligence has “a white guy problem.”³⁶⁵ It is not surprising that advanced technologies tend to be designed by those in power, often reinforcing – albeit inadvertently – existing power structures. However, this has serious implications for creating conditions

³⁶¹ Upturn, As Insurers Embrace Big Data, Fewer Risks Are Shared. *Civil Rights, Big Data, and Our Algorithmic Future*, Sept 2014. <https://bigdata.fairness.io/insurance/>

³⁶² Rosenblat, A. and Randhava, R/ and Boyd, d. and Gangadharan, S.and Yu, C., Data & Civil Rights: Consumer Finance Primer (October 30, 2014). Data & Civil Rights Conference, October 2014. <http://dx.doi.org/10.2139/ssrn.2541870>

³⁶³ Barocas, S. and Selbst, A.D. (2016). Big Data’s Disparate Impact. *California Law Review*, 104, 671-732. doi: 10.15779/Z38BG31

³⁶⁴ Onuoha, M. The Point of Collection. Data & Society Points. February 10, 2016. <https://points.datasociety.net/the-point-of-collection-8ee44ad7c2fa>

³⁶⁵ Crawford, K., “Artificial Intelligence’s White Guy Problem,” *New York Times*, June 25, 2016. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=0

in which some groups will be more privileged than others. For instance, elite or homogeneous teams are likely to inadvertently design systems that advantage individuals like themselves. If we wish to create systems that benefit a diversity of people, then those systems need to be built from a diversity of perspectives.

AI technologies will be used by and will affect everyone in society. It is thus an important first step to take measures to increase the diversity of those who get to design AI systems. This includes increasing diversity among computer scientists and engineers, but also creating new processes through which systems are assessed from a multi-stakeholder perspective, and design processes are required to involve input from those who will be impacted by the application of the AI system.

- AI is built with existing data

Another vector through which to think about how AI may advantage some members of society more than others is to consider the datasets upon which AI is built.³⁶⁶ The current machine learning-driven advances in AI are based on the vast quantities of data and processing power that has developed in recent years. In other words, the “smartness” of AI comes from a system’s ability to process and analyze huge amounts of data, beyond the scale of any individual human.

Turning attention to the datasets underlying AI becomes a way to mitigate the perpetuation of social inequalities or structural disadvantages that may be embodied in data if they are not adequately accounted for.³⁶⁷ Just because a decision is reached by a computer rather than a human does not mean it is free from bias. Data sets must be understood not as self-evident facts, but rather results of particular endeavors shaped by their cultural and historical origins.³⁶⁸

Moreover, it is important to ask: who owns the data? Large U.S. companies like Google, Facebook, Microsoft, Amazon, and IBM are leading the development of AI in large part because it is these companies who have access to and control immense datasets and computing power. Concentrated but privatized data held by corporations will likely advantage those companies already in power, foreclosing opportunities for other market entrants or diverse innovations of

³⁶⁶ Ramirez, E., Brill, J., Ohlhausen, M.K. , & McSweeney, T. (2016). *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*. Report prepared for Federal Trade Commission, United States of America.

https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106_big-data-rpt.pdf

³⁶⁷ Boyd, d. & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society* 15(5): 662-679. doi: [10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878)

³⁶⁸ O’Neil, C. (2016). *Weapons of math destruction: How Big Data increases inequality and threatens democracy*. New York: Crown.

applications. Companies, as well as governments, should work toward creating open and fair data standard practices.

Ethics

5. What are the ethical implications of the development use of artificial intelligence? How can any negative implications be resolved?

Just as in the case of AI's impacts on society, questions of ethical design and use should be understood as a necessary and ongoing point of inquiry, not an area that can be definitively solved. From our perspective, addressing the ethical implications of AI poses a dilemma because questions of ethics are about processing and evaluating risks and benefits or acceptable trade-offs in specific circumstances. The area of ethics should not be thought of as prescriptive, but rather as requiring processes for assessing multiple perspectives and outcomes. In this regard, the current initiatives around developing codes of ethics for AI are an important step on what must necessarily be a multi-step process.

The ethical implications of AI can also be considered from the perspective of how AI has the potential to destabilize existing ethical norms or standards. A clear example is in the field of medicine. Developing and assuring ethical practice is a substantial part of the training and certification of medical professionals. However, when medical AI technologies, which promise capacities of decision-making and evaluation, are developed by engineers and technologists, the standards and priorities of ethical reflection at the heart of medical practice may be lost. In the case of increasing use of AI technologies in healthcare, it may be necessary to develop an expanded notion of ethical responsibility that involves engineers and computer scientists. It is important that AI technologies – and those who design, produce, and implement them—create processes to integrate existing professional ethical practices and norms into AI design and engineering contexts.

Closing statements

AI technologies hold great promises to advance society and address existing problems. However, the potential benefits must not obscure the potential perils of these technologies. These perils have nothing to do with “killer robots” or the coming of “robot overlords.” These perils will be found in the everyday structuring potentials of AI that will benefit some members of society, and leave many others behind. AI technologies are at a crossroads, and now is the time to lay the foundation for the beneficial and just integration of these technologies into society.

M.C. Elish

6. Note

This document has been prepared for the U.K. House of Lords, Select Committee on Artificial Intelligence.

13 September 2017

Mr Graeme Davis – Written evidence (AIC0054)

Date: 3rd September 2017

Artificial Intelligence

The Current state of artificial intelligence

The current state is a race to make the best artificial intelligence with practical uses. This is more limited than the widespread public perception, but improves areas of society that people don't even consider would use AI.

The pace of technological change and the development of artificial intelligence

The current pace is rapid and exciting and should/will be embraced by people from many sectors of the community

The impact of artificial intelligence on society

The impact of AI on society is large and varied. It can lead to improved safety as human error is taken out of some decision making and hard graft tasks like mining, rescue missions etc.

It potentially could reduce jobs as a negative as more mundane tasks are taken over by robots see article (<https://diginomica.com/cdn.ampproject.org/c/diginomica.com/2017/08/29/ocado-put-robots-in-its-warehouse-heres-what-happened-next/amp/>). This could lead to large unskilled parts of the workforce redundant and needing other types of employment (this was originally said in the 1960's with increased technology). However the positive side of this could be to free up time and manpower used to enrich society in other ways.

AI could be used to improve health through diagnosis and treatment of disease, as well as predicting when things could be going wrong before its too late. Predictions of disease outbreaks e.g. malaria?

AI can be used to model and predict problems in architecture, conservation, health, war, agriculture (climate, pest control, fertiliser/pesticide application rates etc.) and even things like weather, landslides etc. (in this case possibly saving lives)

Transportation is an area where AI is going to make a more immediate representation. Self drive cars and lorries are already being tested. This could have far reaching repercussions in both positive and negative ways and will need serious thought in to how it is regulated. Self drive cars for instance seem to lack

some of the human decision making qualities, like when avoiding collisions, if there is a choice between hitting one object or another...will the computer make the right decision? Can AI cars adjust for odd weather related problems like rock falls, mud slides, burst rivers across roadsides etc.? Going forward, who is responsible if the cars are making decisions and an accident does occur? Can't blame the driver, as the AI has made that decision. People who previously couldn't drive, will they now be able to drive e.g. disabled, visual impairments, or people that currently can't pass a test. Would a licence in the future be one in programming a car?

With lorry convoys controlled by AI, would the increased use of this create problems on the roads, like a convoy driving closer and closer together to cut costs causing problems for cars turning off at a junction? (Again, would this lead to another area of jobs being lost?)

The biggest concern for increased AI is that it could be hacked, or corrupted by the wrong people for other means. This could be used in war, insurance claims, or just by general troublesome hackers. Personal information could be gathered from AI and used by others.

The public perception of artificial intelligence

The public perception of AI when it is mentioned is very Sci-Fi based about robots becoming self determined and taking over the world. They very rarely relate it to things actually happening on the ground like warehouse robots, self driven cars etc.

The other perception is that such technology is increasingly taking away human interaction especially when we think of things like social media.

The sectors most and least likely to benefit from artificial intelligence

Most:

Mining, architecture, banking, agriculture, health, elderly, disabled, transportation, space exploration (using AI to travel and record and explore for us)

Least:

Arts, Elderly if not included, sport and environment depending on which areas your talking about

The data-based monopolies of some large corporations

Large corporations can use AI to their own advantage both to the benefits and

disadvantage of people. They can collect your data through things like Amazon, Facebook, Twitter, Google etc. and target things towards you that you may not need, but are related to past purchases, or likes of particular subjects.

An area that is increasingly using the above AI technology to target consumers is gambling. It appears on almost all media feeds and can lead to people being 'sucked' in and making the leap from occasional social gamblers to people with a real problem.

Through the above, ideologies of some corporations can be targeted at people silo-ing them in a particular direction. In the current climate this could lead to people having more right wing or left wing opinions through targeted media from online newspapers, social media etc. Someone who likes a right wing article will get more and more right wing articles appear on their social media feeds leading to a kind of online radicalization done by the computer algorithms.

In the banking world credit scoring decisions and therefore loans, mortgagees etc. are often made by computer decisions without actually having that face to face interaction. This could lead to smaller businesses or people not getting access to the funds they require, despite being reputable.

The ethical implications of artificial intelligence

To make sure lives are not detrimentally effected by AI.

To make sure society does not suffer in terms of employment from increased AI, and that there is something else for workers to fall back on.

To use AI to help other countries in disease, climate and rescue missions.

To make sure that the human element of decision making that will negatively/positively affect someone's life if not completely taken away.

The role of government

The government should be there to ensure that AI is used for the benefit of society. That technology does not end up in the wrong hands, and or is hacked. To ensure that new legislation is made to cover areas like driving licences, how insurance is effected with computer made decisions and hacking. With regards to AI being used in things like war and banking, to insure that the decisions that affect people's lives are not just AI led.

As well as this government should make sure that increased technology does not lead to the further detachment and breakdown of communities and community spirit.

The work of other countries or international organisations

Other countries and organisation should have agreements in where AI can and

should be used, making sure that it is not used where lives can be at threat from such technology. On the positive side the technology can be shared and used for international aid and rescue missions, disease response and prediction and in the aftermath of wars

3 September 2017

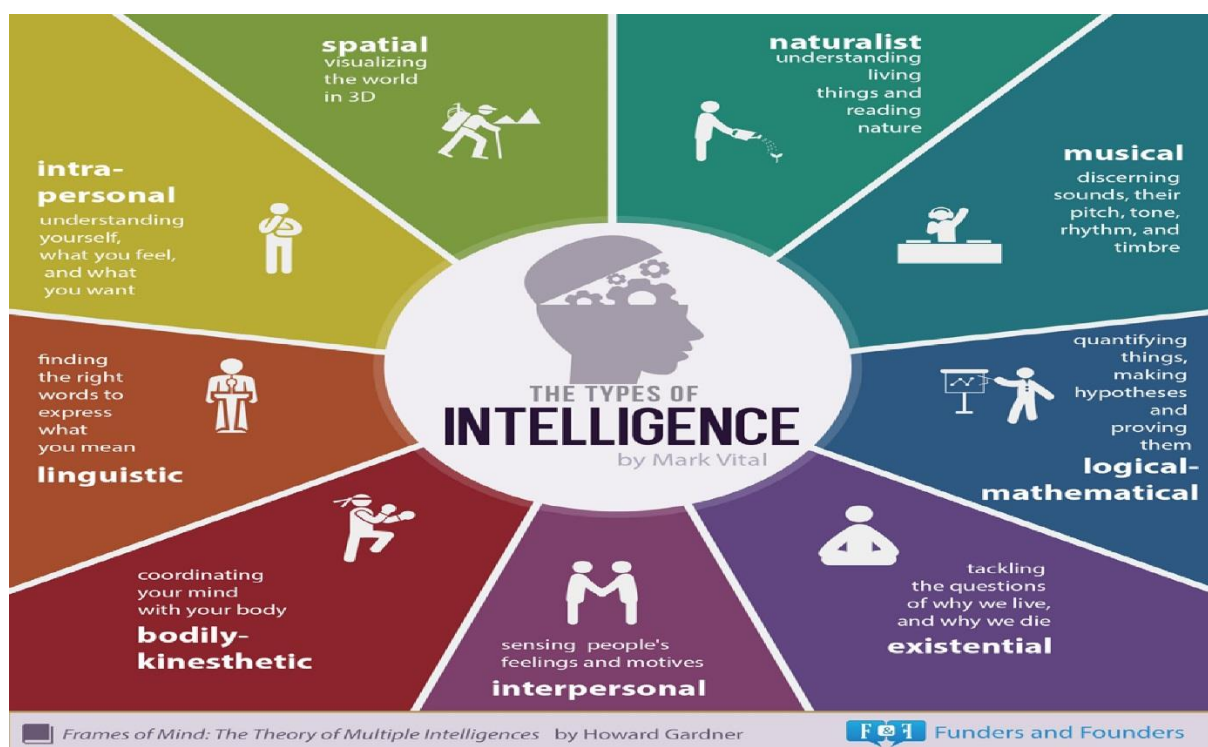
Deep Learning Partnership – Written evidence (AIC0027)

House of Lords - Select Committee on Artificial Intelligence - Call for evidence

What is intelligence?

0. The first thing to do before we start any discussion on artificial intelligence is to carefully define what we mean by intelligence. I offer the following:

Howard Gardner, a Professor of Developmental Psychology at Harvard University, has identified and described at least nine types of intelligence in his theory of multiple intelligences.³⁶⁹ They are described in his 1983 book, *Frames of Mind: The Theory of Multiple Intelligences*³⁷⁰, and are shown in Figure 1 below. They are logical-mathematical, linguistic, interpersonal (social intelligence), intrapersonal (emotional intelligence), existential (spiritual), spatial, musical, bodily-kinesthetic and naturalist. So a truly intelligent system of human-level, general intelligence, will need to incorporate all nine types of intelligence. (IQ tests typically only account for linguistic, logical, and spatial abilities).



³⁶⁹ MI Oasis <http://multipleintelligencesoasis.org>

³⁷⁰ Gardner, Howard, *Frames of Mind: The Theory of Multiple Intelligences*, 3rd ed, Basic Books, 2011

Figure 1 – The various types of intelligence

What about the word “artificial” in artificial intelligence? Artificial here just means anything non-biological. In fact, Zoubin Ghahramani, Professor of Information Engineering at Cambridge University and Co-Director at Uber AI Labs has stated, “The term artificial intelligence is somewhat nonsensical. Something is either intelligent or it isn’t. Just as something either flies or it doesn’t. We don’t talk about artificial flying”. So, artificial in this context simply means anything non-biological. Now that we have understood what intelligence is, and what makes a system intelligent, we can proceed to answer the questions posed.

The pace of technological change

1a. What is the current state of artificial intelligence and what factors have contributed to this?

Since the use of multicore GPU processors in artificial intelligence R&D around 2012, the field of AI development has been accelerating exponentially, driven by the exponential increase in the number of cores available on a single GPU processor – currently up to several thousand cores on the Nvidia V100 and the AMD Vega (see Figure 1). Compare this to a dozen or so cores on the newest CPUs. We have seen a 100-fold speedup in information processing in the last five years due to GPUs, and are on target to see another 100-fold speedup by 2020 due to the development of ASICs optimized for deep learning. Here I use deep learning, artificial intelligence and artificial neural networks (ANNs) interchangeably as they are essentially the same thing. Also, ASIC stands for Application Specific Integrated Circuit and they are, as the name suggests, processors designed to perform one specific application optimally well. Examples of ASICs on or about to come onto the market include Google TPU³⁷¹, Graphcore IPU³⁷² and the Intel Nervana processors³⁷³. All have specified a potential 100X speedup over GPUs. That’s a 10,000X speedup in under ten years, far outpacing the exponential growth of Moore’s Law for CPUs.

³⁷¹ Google TPU <https://cloud.google.com/tpu/>

³⁷² Graphcore IPU <https://www.graphcore.ai>

³⁷³ Intel Nervana <https://www.intelnervana.com>

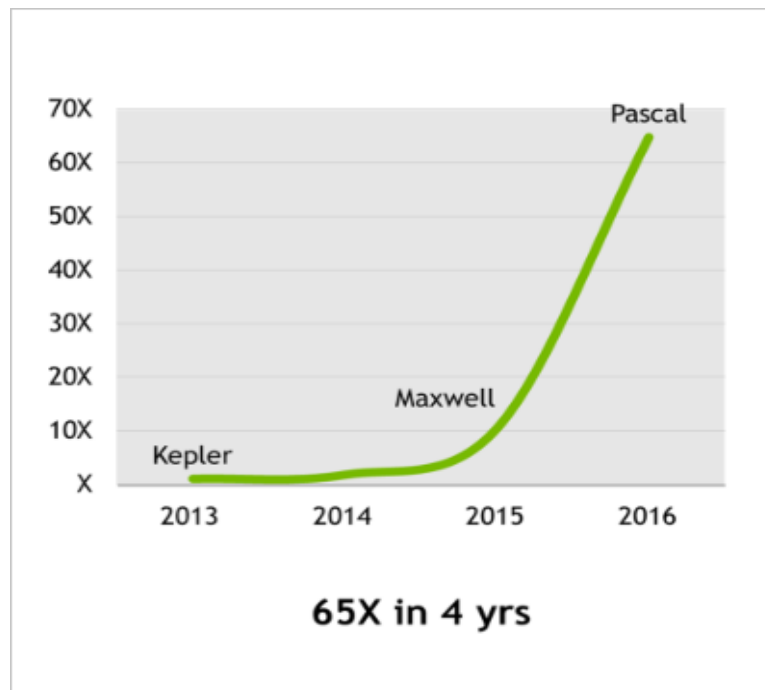


Figure 2 – Nvidia GPU speedup since 2012

Looking slightly further out (five-ten years) we have two new technologies coming to market that will be absolute game changers in the field of information processing in general, and AI in particular. That is neuromorphic computing and quantum computing. Spinnaker³⁷⁴ and TrueNorth³⁷⁵ are examples of neuromorphic processors currently under development in the Human Brain Project and IBM, respectively. A comprehensive report on neuromorphic processors has recently been published in May 2017³⁷⁶. DWave³⁷⁷ have recently brought to market a 2000 core quantum processor. All of these processors can and will run in the cloud, so accessibility, scale and cost are all removed as barriers to entry in AI development.

So rapidly science fiction is turning into non-fiction. The time, therefore, to act is now. Further details can be found in the author's presentation on recent AI

³⁷⁴ Spinnaker <https://www.humanbrainproject.eu/en/silicon-brains/>

³⁷⁵ TrueNorth <http://www.research.ibm.com/articles/brain-chip.shtml>

³⁷⁶ Schuman, C. et al, *A Survey of Neuromorphic Computing and Neural Networks in Hardware*, May 19, 2017 <https://arxiv.org/abs/1705.06963>

³⁷⁷ DWave <https://www.dwavesys.com>

developments³⁷⁸, and in his blog³⁷⁹. An optimistic view of the future, one where our intelligence will merge with machine intelligence, is discussed in some detail in this book³⁸⁰. Right now, AI algorithms can outperform humans in image classification, speech translation and can perform at around human level at simple language generation, mathematical theorem proving, art production and music generation. Over time it is anticipated by AI experts that AI will outperform humans in all areas of intelligence, with many thinking this will occur in the coming decade.³⁸¹

1b. How is AI likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

There are multiple factors involved in the progress of AI, including hardware, software, data, number of people involved and available financing. The above improvement (10,000X in ten years) is coming about through improvements in hardware. Much work is also being done in understanding and improving the algorithms involved in intelligence - learning, reasoning, planning and navigation. Companies such as Deepmind based in London, Google, Apple, Microsoft, IBM, Amazon and Baidu, as well as a myriad of startups all have many, in some cases hundreds of, engineers working on "solving intelligence". Both labelled and unlabelled data sets are increasing in size and quality daily due to the amount of data being generated by such online sites as Google Search, Facebook, YouTube and Instagram. The number of people working in AI is also increasing rapidly as indicated by the headcount within companies, university enrolment and AI conferences such as NIPS and ICML. Also, investment in AI startups is increasing year on year³⁸². The only hindrance is lack of talent and, as stated above, we should be at general artificial superintelligence (ASI) level within the next ten years, with gains continuing to increase exponentially after this.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

From the above overview, I believe the answer is a definite yes.

Impact on society

³⁷⁸ Morgan, Peter, *AI Developments*, Aug 2017, <https://www.slideshare.net/pedronius/ai-developments-aug-2017-v010>

³⁷⁹ Deep Learning Partnership Blogs <http://www.deeplp.com/news.html>

³⁸⁰ Kurzweil, Ray, *How to Create a Mind*, Penguin Books, 2013

³⁸¹ Ibid.

³⁸² Morgan, Peter, *AI Developments*, Aug 2017, <https://www.slideshare.net/pedronius/ai-developments-aug-2017-v010>

3. How can the general public best be prepared for more widespread use of artificial intelligence?

Clearly the advent of general superintelligence (above human level intelligence) is going to bring massive disruption to every aspect of our lives and will have a profound impact on society. Not the least is going to be major technological unemployment brought about by the automation of not only blue collar, but also of white collar jobs – doctors, lawyers, decision makers, etc.³⁸³ In fact, one of the most obvious use cases of AI is in aiding in decisions at the government level - local, national and international. ASI can process petabytes of data much more accurately and quickly than any human, or team of humans, could ever hope to and hopefully, given that they are data driven, make better policy decisions as well. Using ASI to solve social, economic and political issues will hopefully lead to a better and more equitable world with less suffering. A redistribution of wealth will clearly be needed with 50% or more of the population replaced by AI.

A universal basic income has been proposed by many people and think tanks over the years as a solution to the predicted growth in economic inequality, and trials in Canada and other places have proved very successful with positive societal impacts being observed in all of them. A good source of information on this are the many facebook groups devoted to the issue of technological unemployment and basic income.³⁸⁴ There is also some fear around the existential risk of ASI running amok and, intentionally or otherwise, destroying humanity. I personally don't think ASI will be malevolent, in the same way that we humans are not malevolent towards cats and dogs. The issue of AI safety is discussed extensively in the facebook book group AI Safety³⁸⁵ started by the author.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Currently it is the so-called hyperscale Internet companies such as Google, Facebook, Amazon, Microsoft and Baidu that are gaining the most from the development and use of artificial intelligence and data. They have the most data captured for one, and they are leading the curve in terms of AI development and deployment throughout their businesses. In fact, Google has over 4000 instances where they are using AI as highlighted by Head of Google Brain Jeff Dean at a recent Y Combinator talk - see Figure 3 below. The way to address this disparity

³⁸³ Ford, Martin, *Rise of the Robots: Technology and the Threat of a Jobless Future*, Basic Books, 2015

³⁸⁴ Universal Basic Income for Europe

<https://www.facebook.com/groups/basicincomeeurope/>

³⁸⁵ Facebook AI Safety Group <https://www.facebook.com/groups/467062423469736/>

in data and subsequent revenues, is by redistribution of wealth through a taxation system (universal basic income was mentioned above), or perhaps a new economic model, such as a resource based economy.³⁸⁶

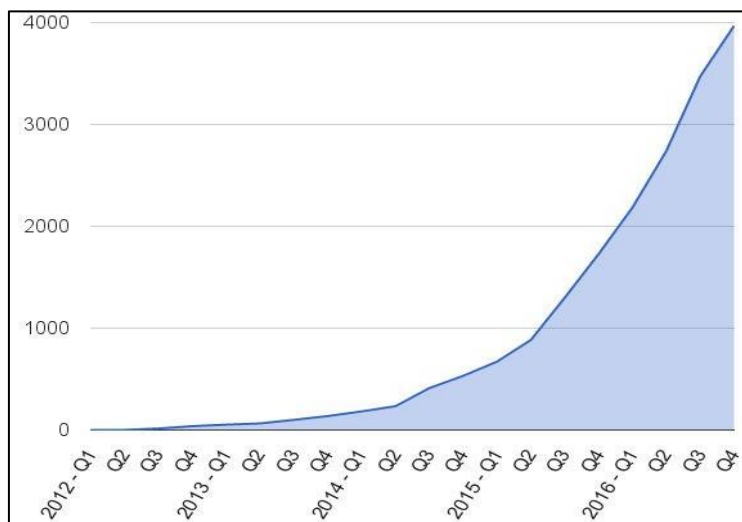


Figure 3 – Growth of Deep Learning at Google

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

Absolutely. The best way is through careful and thoughtful education of what AI is and is not, the limitations on AI development, and the opportunities and the risks that are involved with designing and building superintelligent machines. Use cases include consumer (e.g., social robots), business, medicine, healthcare, drug discovery, human longevity, better decision making, even space exploration. Risks include accidental injuries and deaths due to bugs in the software (a driver was killed in a Tesla self-driving car), poorly trained AI (like the Microsoft chatbot Tay), or hardware failures. There has not been a single technology where people have not been injured or killed in the development or use of it, but we develop and adopt these technologies because the overall benefits and lives saved vastly outnumber any number of casualties.

Industry

³⁸⁶ The Venus Project <https://www.thevenusproject.com>

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

All sectors stand to benefit from the development and use of artificial intelligence. No person or organization will not be impacted by this technology. After all, who can't do with a little more intelligence? This is why it is being referred to as the fourth industrial revolution³⁸⁷, our final invention³⁸⁸ and the new electricity.³⁸⁹ It will become completely ubiquitous, embedded into the fabric of our lives.

7. How can the data-based monopolies of some large corporations, and the 'winner- takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

As stated in question 5 above, the way to address any economic disparity caused by the production of AI is by redistribution of wealth through a taxation system (e.g., universal basic income), or by a new economic model, such as a resource based economy. Legislation, regulation, oversight and enforcement is going to be needed in order to monitor data privacy and to ensure profits made from the use of an entity's private data is equitably shared back to that entity, whether it be an individual or an organization. To ensure AI contributes to the public good and a well-functioning economy, committees and organizations are going to be needed to be set up to monitor and enforce data privacy as well as AI safety at the local, national and international levels. The design, development and deployment of AI is going to need to be regulated by neutral, third party government controlled bodies.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

A strong code of ethics will need to be put in place so that the companies who stand to gain from their lead in AI technology development do not take unfair advantage of their position. This may come in the form of violation of human rights such as privacy, and any kind of unfair exploitation, either monetary or otherwise. Extensions to the current Data Protection Act may be needed, for

³⁸⁷ The Fourth Industrial Revolution by Klaus Schwab
<https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab>

³⁸⁸ Barrat, James, *Our Final Invention*, St. Martin's Griffin, 2014

³⁸⁹ Artificial Intelligence is the New Electricity—Andrew Ng
<https://medium.com/@Synced/artificial-intelligence-is-the-new-electricity-andrew-ng-cc132ea6264>

example, or entirely new legislation will need to be implemented sooner rather than later due to the current acceleration of AI progress.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

Transparency and accountability should, in the author's opinion, be the two foundations on which any legal framework involving AI should be built. Rules under this framework will need policing to deter and penalize violators. Black-boxing may be permissible when national security or personal safety is at stake. A paper outlining some of the technical aspects of black-boxing and verifiability is given in.³⁹⁰

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

The primary role of government is to provide security for its citizens. It is therefore government's responsibility that adequate laws are in place and enforced in order to ensure that all AI development is safe, that certain safety standards are met and upheld in any AI products brought to market, and that the developers are held accountable in the event of accident, negligence and/or criminal activity.

The Government should get involved with all of the companies currently developing AI to understand what it is they are developing and the impacts that their products are having and are going to have on society as a whole. This is extremely urgent and imperative. To do nothing would be irresponsible. Government working with experts in the field and companies developing AI products and services must come up with a framework for overseeing AI safety. Partnership on AI³⁹¹ is an AI safety organization set up by companies but the companies themselves need oversight and regulation. An analogy could be the Atomic Energy Commission or the International Atomic Energy Agency, that had/have oversight on the nuclear industry, both civilian and military.

Learning from others

³⁹⁰ Yampolskiy, Roman, *Verifier Theory and Unverifiability*,
<https://arxiv.org/abs/1609.00331>

³⁹¹ Partnership on AI <https://www.partnershiponai.org>

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

We can learn from current policies around data and protection such as the Data Protection Act and the various safety standards bodies in place that protect the consumer from risks from various products and services they may use or purchase. Consumer Electronics labelling and the ISO organization come to mind. Similar standards committees and standards will need to be created for AI products and services including for intelligent software and robotics.

About the Author: I am CEO of an Artificial Intelligence Consultancy, Deep Learning Partnership. I have been researching in and working in the field of AI for about five years now. I have been approached by two publishers to write books on AI – one technical and one more business focussed. I am working on one at the moment. I also organize a popular Artificial Intelligence Meetup, London Deep Learning Lab with around a thousand members and growing: <https://www.meetup.com/Deep-Learning-Lab/>. Prior to this I worked in the computer networking industry for about ten years, and prior to this I was enrolled in the PhD in physics programme at UMass. I have a strong technical background combined with solid business experience in the tech industry. My profile can be found here: <https://www.linkedin.com/in/peter-morgan-8b7ba2/>.

25 August 2017

Deep Science Ventures – Written evidence (AIC0167)

Submission from:

Mark Hammond, PhD Neuropharmacology & AI
Director at Deep Science Ventures

Deep Science Ventures is a venture builder and pre-seed investor focused on generating highly impactful companies at the interface of science and digital. Originating out of Imperial College's investment house our aim is to flip the traditional research push model and massively scale up the number of deep-tech driven companies addressing major global challenges.

Summary

In summary, we expect to see a 5-10-year uptick in productivity with huge technical advancements before an extended period of wealth consolidation, unemployment and unrest. Relatively minor changes to funding, immigration and data sharing can keep the UK competitive in the near-term but long term adaption will require a completely different model of education and societal worth.

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Current state of artificial intelligence

Basic pattern matching algorithms are being used to replace either repetitive tasks or areas where the amount of data would have been unmanageable for a human to comprehend. Examples include; in finance looking for fraudulent activity by finding the exception to the global norm, in recruitment automating search across millions of profiles, in resource discovery looking for subtle patterns in data that gives away signs of hidden resources in the landscape, in pharma looking to predict drug structure or targets via genomic sequences of word association across papers, in customer service helping to guide to customer to a quicker solution based on the rating of previous interactions, in automation of driving, and in medicine analysing scans and patient records for signs that may be missed by a single doctor.

Developments in the area are largely driven by the top universities, several not for profits set up by the tech elite and big tech companies, particularly Google Deepmind. So far they have chosen to open source both the algorithms and provide access to the computing power to run the algorithms. This has meant that the tech is essentially commoditised and the race for most companies isn't

to develop better algorithms but instead to gather data within a particular domain on which to apply the algorithms to seek a competitive advantage.

Contributing factors

Briefly; AI was previously limited to very simple networks due to the lack of a way to adjust the weights across the network which roughly approximate the connections between cells in our brains. This was solved by a US academic using an optimisation algorithm referred to as 'back propagation' that works across any form of network. This opened the floodgates and created a virtuous circle of advancements in the computing capacity to build larger and more complex networks (largely due to the repurposing of Graphic Processing Units - GPUs), an increased focus on collecting more data to feed the networks and research and venture money into the space.

How we see the space evolving over the next 5-20 years

Phase: 1 ~5 years. Pattern matching

In general the current breed of algorithms are essentially searching for patterns between inputs and outputs in large datasets, which could be anything from experiments in R&D driven companies to stock market movements. These are used to augment the practitioner, giving them a wider and more sensitive view and freeing up their time to do more strategic work, thus creating a more interesting work environment and better bottom line results. The current limitation is on digitising the existing data, not on the capability of algorithms in most cases and both start-ups and big businesses are rapidly chipping away at different bits of the stack within large processes driven organisations. This is of course in addition to the more obvious shift in low skill jobs such as driving and factory work.

Phase 2: ~ 10 years. Integration, strategy, creativity and negotiation

Initially the pattern matching will exist in isolation, for example the risk department in a bank or radiology in a hospital. The next phase will be integration across organisations, and possibly across the economy, via some type of marketplace dynamics. It's hard to imagine algorithms negotiating or generating original content but it has already been demonstrated that the latest machine learning can beat the best human players at both Go and Poker, both of which are highly strategic games, as well as create original content in both an abstract and business context. It's a small step to then integrate these algorithms into business processes, initially as augmentation and ultimately as a replacement to human labor.

Phase 3: ~20 years. Code evolution

Up to this point progress is relatively slow as start-ups and the big tech companies chip away at the opportunities described above, small teams of highly educated machine learning experts drawing on the latest advances in academia. However, we are already starting to see self-written code so we may quickly find that even programming skills are no longer an advantage as machines evolve new code to solve problems and can grasp the requirements from natural language input.

Phase 4: 20 years and beyond

I sit firmly in the camp that there is nothing fundamentally special about our brains, it is merely a complex set of connections joining inputs and outputs. The recent examples of algorithmic negotiation, creativity and strategic planning support this, however more importantly we do know that our brains are severely limited in terms of working capacity (e.g. we can only work on up to 7 pieces of information at one time and in three dimensions) and computers have none of these limitations. Another huge shift will also occur once quantum is able to scale to hundreds of qubits and rapidly optimise extremely complex networks. This moves us into the realm of almost incomprehensible speed of progress, to give this some context just yesterday I sat in on a talk from a major chemical company where they used a computation in 43 dimensions to reduce the time to make a core commodity chemical by 10x. That's faster progress in *1 day than they had made in 50 years* with thousands of people working on the problem. The exponential shift in productivity is almost incomprehensible once these kind of techniques are adopted widely.

2. Is the current level of excitement which surrounds artificial intelligence warranted? Impact on society

Yes. At the fundamental level AI is able to take a unit of work previously done by an individual and turn it into an almost free and instant resource. Initially in the form of isolated pockets and increasingly integrated across organisations and society. In the short term this will massively increase productivity across nearly all industries, create new almost unimaginable products, solve a large part of many of our major challenges such as crop production, resource allocation and health care but it will drive massive unemployment, consolidation of wealth and ultimately erode competitiveness at both a company and global level.

3. How can the general public best be prepared for more widespread use of artificial intelligence? In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated? Public perception

In the 10-20 year horizon people with a growth mindset, technical, creative or human centered abilities will excel as the world shifts to become an orchestra or algorithms led by a small number of conductors. People with a fixed mindset used to processing work will suffer and there will be a 20%+ increase in unemployment. I don't believe new roles will emerge at the same pace as other roles are removed. We're already seeing a raft of early retirement as departments are downsized and this will accelerate. This will increasingly segment society into capital holders, technical elite and everyone else, further driving the current political split and ultimately leading to severe civil unrest.

The typical answer is 're-skilling' but honestly I think the world is moving too fast and too few people are equipped to adapt at the same speed, we probably need to accept that many people are going to be left behind and find ways to rejig society so that value isn't so closely associated with job status. This need to start with schools and a more human centered problem solving based approach to education.

7. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy

Long-standing companies should have a huge advantage in AI as they have years of data. However, they also know that the best talent doesn't want to work for them, they want to build a company that serves multiple customers and because of this most large organisations are striking partnerships to share data. For example we recently built a company that speeds up antibody development (around 50% of drugs are antibodies) from years to minutes, they are partnering with a major pharma company who is providing huge amounts of data with no restrictions under an unspecified deal to be discussed at a later date. Similar deals are occurring across industries from manufacturing to insurance. However, ultimately these companies will be acquired by one of the big existing players (IPOs are increasingly rare) so there will inevitably still be a consolidation of capabilities, data and wealth.

Making it easier, or possibly compulsory, for certain data to be public (and actually accessible in a useful format), specifically when it relates to an individual or publicly funded research would disaggregate data and increase competition. The research bodies are making big steps in this direction already but healthcare and the the commercial world lags far behind. Over a longer term view I would be surprised if the ecosystem doesn't naturally evolve towards a data marketplace of sorts where data is traded according to value, ranging from

anonymised clinical trial data (currently virtually all negative results are hidden!) to individual health records owned and controlled not by the provider but by the individual.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

Personally, I feel that this problem is over stated, humans often have no idea why they made a decision or worse were lead by factors that they were completely unaware of or unwilling to admit to (just look at every crash so far and the continuous problem of medical error as the third leading cause of death). As long as an algorithm is able to significantly outperform a human over an extended period of time and range of input (as relevant to that field) that should be sufficient to give comfort on its capability. It would however clearly be beneficial to performance to understand why a certain choice was made and research is ongoing in this area and this will likely be standard in a few years. A more worrying trend is algorithms running off against each other as had already led to several flash crashes. One way to address is adversarial networks that keep others in check, i.e. one is optimised for making the most money but it has to listen to another that is optimised to minimise market risk.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

From a more near term perspective it would benefit the UK to encourage companies to be less concerned with short term stock market results and invest heavily in AI and acquisitions in the space whilst giving space for workers to up-skill where feasible. The Government could also increase the UK's competitiveness by increasing funding into AI driven research and back truly ambitious start-ups and public-private initiatives that leverage AI. Far too much translational, angel and VC money is focused on 'safe' incremental plays that will rapidly be made irrelevant by Chinese or US competitors who are thinking at a much larger scale.

By far the biggest risk to the UK losing out in this race is the current focus on arbitrary immigration targets separating ourselves from EU wide research initiatives. We are already seeing highly skilled machine learning experts sent back to volatile countries on nonsense administrative issues and many experts resettling in Europe voluntarily due to the anti-EU sentiment. Meanwhile UK academics are getting removed from H2020 applications. The key advantage the UK has is top universities and we need to make it as easy as seamless as possible to attract and keep top talent and keep them in the UK. We cannot compete in this field against the US and China as just the UK, it must be as a European effort regardless of how Brexit plays out.

Personally, I don't believe that regulation of AI as a technology is necessary as long as the current very open approach of the major tech companies continues. The regulation should sit at the application level and be adapted for such a dynamic system. For example, in financial services the kind of adversarial algorithm mentioned above, in healthcare algorithms should have to pass a bar similar to medical exams, in driving they should be as good as a competent driver, in pharma the resulting drug should still have to undergo the same trials as human designed alternatives.

6 September 2017

DeepMind – Written evidence (AIC0234)

Introduction

DeepMind makes this submission to the Committee as part of the Select Committee on AI's call for evidence. We welcome the Select Committee on AI's research into the economic, ethical and social implications of advances in artificial intelligence, and appreciate the opportunity to provide input.

We write with reference to eight of the eleven questions asked in the initial call for evidence and in some cases have grouped questions together. We have also submitted written responses to the questions on health asked during the oral evidence sessions.

Q2. Is the current level of excitement which surrounds artificial intelligence warranted?

(1) The work we're undertaking at DeepMind gives us reason to be optimistic that, over time, breakthroughs in artificial intelligence research will be able to help society tackle some of its toughest problems. We're working on some of the world's most complex and interesting research challenges, with the ultimate goal of building general-purpose learning algorithms that can work across a variety of tasks. To do this, we've developed a new way to organise research that combines the long-term thinking and interdisciplinary collaboration of academia with the relentless energy and focus of the very best technology start-ups, alongside a clear social purpose.

(2) This approach has already led to significant breakthroughs, such as our computer program, AlphaGo, which defeated a professional Go player in a landmark achievement that experts agreed was a decade ahead of its time. During the games, AlphaGo played many highly inventive winning moves, several of which were so surprising they overturned hundreds of years of received Go wisdom³⁹². These moments of algorithmic inspiration give us a glimpse of why AI could be so beneficial for science: the possibility of machine-aided scientific discovery. We believe the techniques underpinning AlphaGo are general-purpose and over time could be applied to a wide range of other domains.

(3) Our DeepMind Applied team already works with experts in different fields to use these techniques to tackle real world challenges. For example, our systems have reduced the amount of energy used for cooling in Google's data centres by up to 40 percent³⁹³, and we're collaborating with clinicians in the UK's National

³⁹²[The story of AlphaGo so far, May 2017](#)

³⁹³[DeepMind AI Reduces Google Data Centre Cooling Bill by 40%, July 2016](#)

Health Service to deliver better care³⁹⁴ for conditions that affect millions of people worldwide.

(4) Ultimately we hope that new scientific breakthroughs, driven by advances in machine learning, can make the crucial difference in helping us to prosper in this increasingly complex world, helping us understand and respond to tough challenges from climate change, to resource scarcity, from curing complex diseases to addressing discrimination, all fields that could otherwise remain intractable.

Q3. and Q5. How can the general public best be prepared for more widespread use of artificial intelligence? Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

(5) Collaboration, diversity of thought, and meaningful public engagement are key if we are to develop and apply artificial intelligence with widespread benefit. At DeepMind, our recently launched research unit, DeepMind Ethics & Society will work with a variety of partners in an effort to engage with and include a broad set of viewpoints.³⁹⁵

(6) For example, in early 2018 we will begin a public lecture series in partnership with the Royal Society, to explore the societal implications of cutting-edge AI research, building on the Society’s recent projects in these areas. We will also work alongside the RSA on a series of citizen juries on the use of AI in criminal justice and democratic debate.³⁹⁶ These events will use immersive scenarios to help participants understand the ethical issues raised by AI, and to facilitate meaningful public engagement on some of the most pressing issues facing society today.

(7) While we will undertake our own research, it is also clear that this is a debate that must extend beyond any one company or sector. DeepMind is a founding board member of the Partnership on AI to Benefit People and Society, whose Board has equal representation from corporations and non-profits. The Partnership on AI was established to study and formulate best practices on AI technologies, to advance the public’s understanding of AI, and to serve as an open platform for discussion of AI and its influences on people and society. The organisation will explore a wide range of research themes, including questions around bias and discrimination in algorithms, safety and robustness of machine learning systems, and the impact of machine learning on automation and labour.³⁹⁷

³⁹⁴[DeepMind Health, February 2017](#)

³⁹⁵[Why we launched DeepMind Ethics & Society, DeepMind, October 2017](#)

³⁹⁶[The role of citizens in developing ethical AI, RSA, October 2017](#)

³⁹⁷[Partnership on AI website, 2017](#)

Q4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

(8) Everyone has a right to participate in debates surrounding systems that have such a profound impact on our daily lives. It's in this collaboration between people and algorithms that incredible scientific and social progress lies over the next few decades. If we can deploy these tools broadly and fairly, fostering an environment in which everyone can benefit from them, we have the opportunity to enrich and advance humanity as a whole.

(9) At DeepMind, we are committed to ensuring that the immense potential impact of these technologies is of overall benefit to society, and that by their very design, they reflect our highest collective selves. We recognise that AI can be disruptive, with uneven and hard-to-predict implications for different affected groups. As scientists and practitioners working in this area, we have a responsibility to support open research and investigation into the wider impacts of our work, in order to secure its safety, accountability, and potential for social good.

(10) Through our new research unit, DeepMind Ethics & Society, we have committed to producing and supporting original, rigorous, interdisciplinary research that can contribute to answering some of these ethical dilemmas. Addressing these critical issues will require informed debate amongst policy makers, the broader policy community, the AI field, and society at large.

(11) In addition to being a founding board member of the Partnership on AI, we have also supported the launch of ten postdoc positions at the AI Now Institute at NYU, an independent, interdisciplinary research initiative dedicated to understanding the social and economic implications of AI.³⁹⁸

Q7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

(12) Machine learning technologies benefit not only from large volumes of data, but also the right types of data for innovation and research. At DeepMind we have made extensive use of simulated environments allowing significant research progress without access to public datasets, and we have made one such environment, DeepMind Lab, available under an open source license.³⁹⁹

³⁹⁸[DeepMind Partners - AI Now Institute, New York University, DeepMind, 2017](#)

³⁹⁹[Open-sourcing DeepMind Lab, DeepMind, December 2016](#)

(13) In many research areas, simulation is difficult or intractable, and so open access to data is needed to enable successful research. We recommend promoting the adoption of open, interoperable data standards that will enable end users to easily transfer their data to a competing service if they wish. In addition, it is critical to ensure that the rights of individuals to privacy and control over their data, and the integrity and security of institutional data are fully respected. We believe that this is best supported by transparency around what data has been used, how, by whom, and with what results, an approach taken by our Verifiable Data Audit project in our healthcare work.⁴⁰⁰

(14) Secure data will be one of the key foundations upon which success in AI research and innovation is built. Managing data securely is critical to being able to continue to apply AI and machine learning to improve the apps and services we all rely upon. As secure and protected ways of providing data continue to evolve, government should play a significant role in supporting academic research into world-leading data security practices, with widespread UK adoption in mind. The UK should also continue to make firm commitments and progress towards a strong and innovative data policy that ensures the highest standards of data portability and security. This should include a continued public commitment to ensuring encryption standards are never weakened, given the vital role such standards play in keeping data safe and secure.

(15) We welcome the recommendations put forth in the recent AI Review that government and industry should deliver a programme to develop Data Trusts in order to stimulate the secure and mutually beneficial exchange of data.⁴⁰¹

Q8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

(16) The development of AI creates important and complex questions. Its impact on society - and on all our lives - is not something that should be left to chance. Beneficial outcomes and protections against harms must be actively fought for and built in from the beginning. But, in a field as complex as AI, this is easier said than done. The Budget announcement to create a new Centre for Data Ethics and Innovation to enable and ensure safe, ethical and ground-breaking innovation in AI and data-driven technologies is welcome.

(17) At DeepMind, we start from the premise that all AI applications should remain under meaningful human control, and be used for socially beneficial purposes. Understanding what this means in practice requires rigorous scientific inquiry into how this is best implemented in the full range of application scenarios.

⁴⁰⁰[Trust, confidence and Verifiable Data Audit, DeepMind, March 2017](#)

⁴⁰¹[Recommendations of the review, Growing the Artificial Intelligence Industry in the UK, Government Publications, October 2017](#)

(18) This is why we recently launched our new research unit, DeepMind Ethics & Society, which will help us explore and understand the real-world impacts of AI. It aims to help technologists put ethics into practice, and to help society anticipate and direct the impact of AI so that it works for the benefit of all. If AI technologies are to serve society, they must be shaped by society's priorities and concerns.⁴⁰² DeepMind Ethics & Society will organise around six key ethical challenges that we believe are facing the field of AI.⁴⁰³

(19) We also believe that to mitigate potential negative implications, it is important that we break from our traditional silos and ensure we are working together as a society. That is another reason that we helped to start the Partnership on AI, a new type of organisation that aims to bring together industry, academia and civil society to conduct research into the potential implications of AI and come up with a set of best practices and standards around its deployment.

Q9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

(20) Both oversight and understanding of AI systems is required to ensure that society benefits from this technology. We believe that it should be possible to provide a meaningful explanation for the decisions and outcomes produced by AI systems, and that these should be open to challenge. This is possible through transparency about the data used to train AI systems and the methodologies applied, without providing detailed information on the algorithms themselves. Equally important is that machine learning researchers develop methodologies for interpreting the behaviour of AI systems, such as DeepMind's research on "virtual brain analytics".⁴⁰⁴ Furthermore, ethical outcomes in the technology sector depend on far more than algorithms and data - they depend on the quality of societal debate and accountability.

(21) At DeepMind, we believe that companies should provide more visibility around data access and use. End-users, service providers, contracting organisations and technical auditors should each be able to understand who accessed their information, for how long and under which policy. Earlier this year DeepMind Health announced the development of the Verifiable Data Audit tool, which will develop a methodology to explain when, how and why a patient's data is used, while minimising the possibility of falsification or omission.⁴⁰⁵ We believe that VDA offers a way to establish real confidence about how data is being used

⁴⁰²[Why we launched DeepMind Ethics & Society, DeepMind, October 2017](#)

⁴⁰³[Key Ethical Challenges, DeepMind Ethics & Society, DeepMind, October 2017](#)

⁴⁰⁴ Hassabis, Kumaran, Summerfield and Botvinick. "[Neuroscience-Inspired Artificial Intelligence.](#)" *Neuron* 95, no. 2, July 2017

⁴⁰⁵[Trust, confidence and Verifiable Data Audit, DeepMind, March 2017](#)

in practice, and could bolster society's trust and confidence in the vast amounts of data powering our most important institutions.

Q10a. What role should the Government take in the development and use of artificial intelligence in the United Kingdom?

(22) We were pleased to see the Government's commitment in the Industrial Strategy White Paper to ensuring that the UK is at the forefront of the AI and data revolution - including funding for PhDs and a new Centre for Data Ethics and Innovation. The advent of new technologies has always helped shape our social and economic landscape, and we should expect that increased use of AI and machine learning will be no different. In many sectors, machine intelligence will augment and enhance the work that people do, enabling them to be more effective in the same roles. As with all technological innovation, we should expect that new areas of economic activity and employment will be made possible, and some types of work and some skills will decrease in relevance.

(23) It is therefore important that government focuses on expanding commitments to education and diversity, research and development, career resilience, and infrastructure. In the long-run, government will of course play the pivotal role in any regulation of AI and standards.

10b. Should artificial intelligence be regulated? If so, how?

(24) According to the Royal Society, if the broad field of artificial intelligence is the science of making machines smart, then machine learning is a technology that allows computers to perform specific tasks intelligently, by learning directly from examples, data, and experience.⁴⁰⁶ Despite many recent breakthroughs in machine learning, the field and its applications are still nascent. As these opportunities are still emerging, we advocate a nuanced approach to regulation that will allow innovative uses to flourish and reach their full potential. Given the global reach of machine learning technology, we believe it sensible to pursue regulatory harmonisation and stakeholder consultation that seeks out international perspectives.

(25) There is still value, however, in thinking early about the potential effects of AI on society and the regulatory responses that may be necessary in future. We believe consensus-driven best practices and innovative governance mechanisms will play an important role in ensuring the flexibility needed to drive growth in this sector, while simultaneously developing robust safeguards. We also encourage investment in rigorous research to support the development of evidence-based policymaking.

⁴⁰⁶Machine learning: the power and promise of computers that learn by example, Royal Society, April 2017

(26) There are also some potential real-world applications of these technologies that deserve early attention, in advance of their widespread development and use. For instance, we are concerned about the possible future role of AI in lethal autonomous weapons systems, and the implications for global stability and conflict reduction. We support a ban by international treaty of lethal autonomous weapons systems that select and attack targets without meaningful human control.⁴⁰⁷ We believe this is the best approach to averting the harmful consequences that would arise from the development and use of such weapons. We recommend the government support all efforts towards such a ban.

Question asked in oral evidence session on AI in Healthcare

Q1. To what extent is AI already used in healthcare? Where in health do you see the biggest potential for the use of AI?

- **What are your impressions of the Government's recent AI Review? What are its implications for healthcare? Does the review go far enough?**

(27) AI⁴⁰⁸ is currently not widely used in healthcare in the NHS or indeed any other health system. For example, the only DeepMind product currently in use for direct care by the NHS – the Streams app at the Royal Free Hospital in North London – does not use any form of artificial intelligence or deep learning technology.⁴⁰⁹ Streams is a secure mobile phone app that aims to address what clinicians call “failure to rescue” – when the right nurse or doctor doesn’t get to the right patient in time. Each year, many thousands of people in UK hospitals die from preventable conditions like sepsis and acute kidney injury⁴¹⁰, because the warning signs aren't picked up and acted on in time. Our goal with Streams is to help the NHS move from pagers and paper lists to modern digital technologies, which we feel is an important stepping stone before we can realise the potential of AI-enabled healthcare.

(28) However, within the NHS, there may be institute-specific examples of machine learning software used in patient care that aren’t registered as medical devices because they are currently only being used at the hospital Trusts that developed the technology. Such examples are, to the best of our knowledge, uncommon. Some examples do have an EU CE-marking, such as Zebra Medical. In January 2017, the US Food and Drug Administration (FDA) gave its first clearance to a technology that leverages deep learning in a clinical setting (to

⁴⁰⁷As currently being discussed by the UN Convention on Certain Conventional Weapons, United Nations, November 2017

⁴⁰⁸By AI, we mean the use of deep learning algorithms that are trained to perform specific tasks by extracting patterns and information from a set of data, without humans programming how to achieve this.

⁴⁰⁹Why doesn't Streams use AI? DeepMind, November 2017

⁴¹⁰ <http://qualitysafety.bmj.com/content/early/2012/07/06/bmjqs-2012-001159>

Arterys, cloud-based medical imaging software). There are also several examples of companies that have developed AI applications for healthcare, and which are in the process of obtaining approval from the FDA to deploy their products in patient care, including Lunit.IO.

(29) Although AI in healthcare is currently uncommon, we do believe that over time it has the potential to make a significant positive impact and we are likely to see important breakthroughs in this area. In the medium term, we believe that AI technology could help clinicians with more accurate analysis, diagnosis and triaging, allowing them to deliver faster treatment to the patients who need it most. In the long term, we think AI tools will be able to learn how to analyse clinical test results and scans to predict whether a patient might be at risk. However, this should only be done with patients placed at the centre of research design, addressing genuine clinical need and mitigating against potential biases. And while we should make sure technologies that offer benefit are implemented as soon as possible, it's important that independent clinical trials are conducted to provide the level of evidence required for safe, effective implementation.

(30) Our own pursuit of positive clinical impact involves two separate strands of work: the immediate development and deployment of our mobile app Streams (which doesn't involve any AI or deep learning) and a set of longer-term AI research projects.⁴¹¹ If any technologies developed as a result of our research become proven, we hope to bring these benefits together with Streams thoughtfully, using a transparent, ethical approach to investigating possible clinical impacts. In this way, we hope patients and clinicians can benefit from AI-supported care wherever they are in the hospital.

Q2. In your experience, how does the public view the use of AI in healthcare? How aware are they of its use? What could be done to improve the public perception of the use of AI in health?

- **If and when a medical AI application goes wrong and, for example, makes a decision or provides advice which adversely affects a patient, how should liability and compensation be handled? Do we need new mechanisms for handling this?**

(31) In our experience, there is a very low public awareness of how technology is used in the NHS, and the way patient data is routinely used in the provision of care.

(32) For this reason, we think that wider public engagement must be a priority. As part of our own transparency efforts, we worked with the late Rosamund Snow, patient editor of the British Medical Journal, to draft DeepMind's first formal patient and public engagement strategy, and to host our first patient and

⁴¹¹DeepMind Health and Research Collaborations, DeepMind 2017

public engagement event in September 2016 at our London offices.⁴¹² Further workshops were held in London and Manchester in July 2017. We are currently building on this engagement by holding additional focus groups for patients and carers in London in December.

(33) We recognise that listening to and learning from patients and the public is an ongoing process, and we look forward to continuing to grow our engagement programme. However, we believe there is also a role for the Government and the NHS in informing patients about how data is used in NHS healthcare, as well as the potential transformational benefits of new technologies such as AI. Only with transparent and open communication will it be possible for people to feel reassured about critical issues such as data security and privacy.

(34) We also welcome innovative attempts at more democratic forms of patient and public engagement, such as the Connected Health Cities citizens' juries, which addressed scenarios that reflect the current realities of NHS data processing.⁴¹³ We have also been in close contact with organisations involved in the debate around patient data, including the Wellcome Trust, and have contributed our experience to their Understanding Patient Data programme, which aims to educate the public about how data is used in medical care.⁴¹⁴

(35) When it comes to liability and compensation, these will be critically important issues if ever artificial intelligence technology were to replace the expert opinion of a medical professional. However, at this juncture it must be noted that the efforts in AI that are currently most likely to lead to use in clinical practice – such as using deep learning to analyse and classify medical images like eye scans much more efficiently than current techniques allow – will not involve replacing an expert human's clinical judgement, but instead augmenting it, with final responsibility for diagnosis and treatment remaining with the clinician.⁴¹⁵

Q3. Should all publicly-generated health data be made publicly available (subject to anonymisation) in order to encourage progress in AI research and innovation? How could this be best and most safely achieved?

- **Would do you think about recent proposals for data trusts that broker safe access and ethical use of data, or the Royal Society and British Academy's proposals for a data stewardship body?**
- **Some have suggested that the National Institute for Health Research should consider setting up an AI BioResource, similar to the approach taken to genomics. Would this be a sensible approach?**

⁴¹²[For patients, DeepMind, February 2017](#)

⁴¹³[Citizens' Juries, Connected Health Cities, November 2016](#)

⁴¹⁴[Understanding Patient Data, 2017](#)

⁴¹⁵[DeepMind and Moorfields Eye Hospital NHS Foundation Trust, DeepMind, May 2017](#)

(36) There are complex questions for the NHS to consider about how it can best use the data it holds for innovation or research in the future, and we believe that patients, clinicians and the public should be part of the conversation about such matters. There are potential benefits to making some NHS data publicly available – for example, Trusts being able to independently demonstrate that clinical machine-learning models from third parties work as intended, or helping to develop new research into AI that could diagnose specific health conditions. However, if NHS data is being made publicly available for research, there must be robust governance and consent measures in which both patients and companies can have confidence.

(37) The proposed AI BioResource is a good example of the kind of framework that we believe would be productive at ensuring fair, transparent and productive access to research data, encouraging innovation, whilst also protecting patient privacy and preventing data misuse through a robust governance framework. We also believe that the National Institute for Health Research has an excellent track record of encouraging research in the NHS while also upholding high standards of governance.⁴¹⁶ This would support the principle of open and transparent access to data for the good of the NHS, whilst enabling necessary oversight.

(38) We also welcome recent proposals to broker safe access to data and its ethical use to foster and encourage innovative research, whilst also ensuring that only approved individuals and organisations can access data, with clearly defined restrictions on the purposes for which data can be used, and for how long it can be retained.

(39) However, it is important to note that there is not currently a bank of NHS-generated health data ready for third parties to use in AI research projects. Much health data is currently not fit for such use unless it is appropriately formatted and curated, which is a significant amount of work, requiring considerable time, resources and expertise. This has been demonstrated through the project to set up the OPTIMAM database for research using breast scans.⁴¹⁷ It is important to flag this because data that is not appropriately prepared is either not suitable for AI research, or could potentially lead to inaccurate results.

Q4. Should the NHS be recompensed or incentivised when it makes data available to companies for the purposes of AI development? If so, how and under what conditions?

- **Should the Government retain the ownership of algorithms developed using NHS data? Why/why not?**
- **In the deal between DeepMind and the Royal Free NHS trust, how was the value of 1.6 million patient records determined, and has**

⁴¹⁶[National Institute for Health Research, 2017](#)

⁴¹⁷[OPTIMAM, Mammography Image Database, Royal Surrey NHS Foundation Trust, 2017](#)

the NHS retained any rights over the applications, such as the Streams app, that have been developed as a result?

(40) The Royal Free has not sold DeepMind any of its patient data and at no point have any algorithms been developed using this data.

(41) The Streams app was developed entirely with synthetic data; no Royal Free patient data was used to develop the app. In creating the first version of Streams, we worked closely with Royal Free clinicians to understand exactly the problems they face in using current NHS IT, and discussed in detail what they'd want from an app like Streams in order to ensure that it would meet their needs. However, patient data was not part of this process, and the technology was built entirely by DeepMind, which is why DeepMind owns the intellectual property of the app. Our NHS partners agree, and this agreement is reflected in our contracts with each of them.

(42) DeepMind acts as a Data Processor for the Royal Free, with the Trust remaining the Data Controller, as defined by the Data Protection Act 1998. Consequently, this data is not and cannot be used for any form of research. DeepMind cannot exploit it for research or product development purposes in any way.

(43) DeepMind's relationship with the Royal Free, and its handling of data, operates on the same legal basis as the many other organisations that NHS hospitals instruct to process patient data to help them provide patient care.

(44) When it comes to AI development, we agree that the NHS should be recompensed when it makes data available to companies for the purposes of AI development. Clearly there are many ways to recognise and return value and we expect that conversation to continue in the months and years ahead.

Q5. Some of our previous witnesses have suggested that data can never be truly anonymised. Is this correct? How could NHS data be used safely and securely for the benefit of society?

- **Is there a level of data anonymisation you believe to be 'good enough' for the purposes of healthcare?**

(45) DeepMind processes two types of health data. For our Streams application, which helps clinicians provide direct medical care to their patients, we process identifiable patient data on behalf of data controllers such as the Royal Free NHS Foundation Trust. Under our agreement with the Royal Free, DeepMind acts as a Data Processor, with the Royal Free remaining the Data Controller. DeepMind cannot use this data for any form of research, and it can only be processed to

provide the Trust with the services set out in our contractual agreement and IPA with them.⁴¹⁸⁴¹⁹

(46) We hold patient data at the very highest levels of security, and ensure that all data is encrypted, logged and strictly governed. Our security systems and processes have undergone and passed multiple NHS audits, and data is stored in an NHS Digital approved data centre located in the UK. Data transmitted between machines is also end-to-end encrypted, and all equipment is physically secured within a locked cage. All backups within our systems are also conducted over secure, encrypted links. All data access is logged and available for audit, and once data is no longer required, we permanently delete it from our systems. Where applicable, we also destroy any encryption keys associated with that data. Any storage device that is retired from service in our data centre is physically destroyed to ensure there is no possibility of data leakage or recovery.

(47) The second type of data we process is de-personalised health data, which is used by DeepMind’s research scientists to explore whether new, innovative AI algorithms can be created that can help predict or detect disease. Our work into optical coherence tomography (OCT) scans with Moorfields Eye Hospital is an example of such a project, where we are exploring whether AI could be used to spot signs of eye diseases that can potentially cause blindness much more effectively than current techniques allow.⁴²⁰

(48) However, we agree with the characterisation of patient data from the Wellcome Trust’s Understanding Patient Data group.⁴²¹ They argue that data does not sit neatly in either “identifiable” or “anonymised” categories, but that there is, in fact, a spectrum of identifiability that spans a wide range of data types between those that are unambiguously personally identifiable, and those that are anonymous.

(49) We believe that no more data should be used than is necessary to accomplish a project’s stated objectives, and that data should be as far away from the identifiable end of the spectrum as possible, while still allowing the research to be conducted. Consequently, we do not think that there is one level of data anonymisation that is “good enough” for all research problems, as the required level of anonymisation can vary on a project-by-project basis.

(50) In addition to de-identification, we believe that for AI research projects, there should be:

⁴¹⁸Services Agreement between DeepMind Technologies Ltd. and Royal Free London NHS Foundation Trust, DeepMind documents, November 2016

⁴¹⁹Information Processing Agreement between DeepMind Technologies Ltd. and Royal Free London NHS Foundation Trust, DeepMind documents, November 2016

⁴²⁰DeepMind and Moorfields Eye Hospital NHS Foundation Trust, DeepMind, May 2017

⁴²¹What does anonymised mean? Understanding Patient Data, 2017

- Technical controls to ensure appropriate data security (such as encryption, installing software updates, and ensuring the system can detect intruders)
- Authentication mechanisms to prevent unauthorized access to data (such as lists of who has the right to access data, strong passwords for those who do, and two-factor authentication like fingerprint login)
- Legal controls on use of and access to data, including in contracts with employees and contractors
- Information governance training for all employees and contractors who have access to data

(51) Finally, we strongly believe that technical measures should be put in place so that data processors can prove, with no possibility of falsification, that anonymised patient data is only being used for approved research purposes. For example, in 2017 we announced a research project called Verifiable Data Audit (VDA), to create a cryptographically verifiable log of all the interactions with a specified data set.⁴²²

(52) We also support the recommendation of the National Data Guardian in her 2016 review of data security, consent and opt-outs that the Government should introduce stronger sanctions, including criminal penalties in the case of deliberate or negligent re-identification, to protect an individual's anonymised data.⁴²³ We believe that such sanctions, combined with verifiable data audit mechanisms described above, could expand the possibilities for more safe and secure use of NHS data for AI research.

Q6. Increasingly, AI applications are being developed in specific contexts, using unrepresentative datasets, before they are sold in places with very different clinical contexts and patient demographics. What assurances should the NHS look for that these systems are appropriate for use in a British context?

- **Do you audit or check the systems you use or develop, to ensure they are as fair and appropriate to the clinical context here as possible? Is this something we should be considering?**
- **What level of transparency should be insisted on when using AI in a clinical context?**

(53) The systems developed by DeepMind in use today by the NHS do not currently use artificial intelligence. However, we are committed to the open and transparent evaluation of the effects of our non-AI system, Streams, on the NHS hospitals in which is deployed. We have therefore published an open-access

⁴²²[Trust, Confidence and Verifiable Data Audit, DeepMind, March 2017](#)

⁴²³[Review of data security, consent and opt-outs, National Data Guardian, Government publications, July 2017](#)

protocol, in conjunction with our academic and clinical partners, describing how its impact will be evaluated.⁴²⁴

(54) As we conduct AI research, separately to our work supporting direct care with Streams, we are keen to ensure that we work with datasets that represent a fair and appropriate mix of patient cases, and that our research does not discriminate or disadvantage any segment of the population on the basis of race, gender, sexual orientation, age, disability or any other protected categories, to ensure their maximal effectiveness if they are used to support patient care in the future.

(55) Unrepresentative datasets can lead to unintentional bias. For example, an algorithm that has never trained on data from men, or from women, or from a certain ethnic group, could misidentify or fail to recognise inputs from these groups at a later date. Representative datasets are important for guarding against the chance of producing a model that does not perform correctly for minority groups that aren't well represented in the dataset.⁴²⁵⁴²⁶

(56) We believe that every effort should be made to validate the applicability of any new algorithm to NHS patients prior to its procurement by the NHS or deployment in clinical practice. While this is ultimately a matter for the MHRA and/or other regulatory bodies, we believe that it would be valuable for NHS hospital trusts that intend to procure an AI model from a third-party provider to hold a dataset that reflects its patient population so that they can use it to validate existing models that are trained elsewhere, just as is currently done for non-ML based risk scores in the NHS.⁴²⁷⁴²⁸

(57) DeepMind is committed to transparency of the operation of our AI models. In our early discussions with clinicians and patients, it became clear that the interpretability of AI models was crucially important if they are to understand how our models work. We are currently working to address this in our research.

(58) In addition, we aim to set new standards for transparency in healthcare. DeepMind has therefore committed to taking a number of proactive steps to enable meaningful scrutiny of our projects, legal agreements and data use.

- *Independent Reviewers:*

⁴²⁴Alistair Connell, Montgomery H, Morris S *et al.* "[Service evaluation of the implementation of a digitally-enabled care pathway for the recognition and management of acute kidney injury](#) [version 2; referees: 2 approved]." *F1000Research*, 6:1033 Last updated: 10 August 2017

⁴²⁵Hardt, Price and Srebro. "[Quality of Opportunity in Supervised Learning](#)." *arXiv:1610.02413*, October 2016

⁴²⁶Angwin, Larson, Mattu and Kirchner, "[Machine Bias](#)," *ProPublica*, May 2016

⁴²⁷Corfield, Gowens, Rooney and Silcock. "[Validation of the National Early Warning Score in the prehospital setting](#)." *Resuscitation* 89, April 2015

⁴²⁸Pirneskoski, Nurmi, Olkkola and Kuisma. "[21 Prehospital national early warning score \(NEWS\) does not predict one day mortality](#)". *BMJ Open* 7, no.3, May 2017

- We've asked a number of respected public figures to scrutinise our information processing agreements, our privacy and security measures, and our product roadmaps in the public interest as unpaid Independent Reviewers of DeepMind Health.⁴²⁹
- *Public contracts:*
 - We have published all our contracts with NHS hospitals with only minor redactions.⁴³⁰
- *Public and patient involvement:*
 - We worked with the late Rosamund Snow, formerly patient editor of the British Medical Journal, to devise an innovative patient and public involvement strategy for our work in the NHS.⁴³¹ Our work to date has included a major patient summit and workshops around the country, with more planned next year. We are also exploring how patients and the public can get involved with DeepMind Health to develop and co-design new healthcare technologies.

Q7. Does the NHS have the capacity to take advantage of the opportunities represented by AI technology, and to minimise the risks? Are the clinicians and other healthcare professionals equipped with the necessary skills to take advantage of AI technology in their practice? What could be done to help them? How do you see a productive and safe co-operation between doctors and AI working?

(59) We believe there is great potential for the NHS to take advantage of the opportunities in AI. The UK is home to some of the leading hospitals and clinical experts in the world, as well as being the home of the most cutting-edge advances in AI. However, in our experience, the NHS currently is not able to set aside resources to explore in full the potential that AI holds, which leaves clinicians and other healthcare professionals ill-equipped to make the most of these opportunities. As the NHS will face increasing pressures from an ageing population with more complex healthcare needs, AI technologies could help alleviate some of these pressures, so we believe that clinicians should be given greater support to understand the potential of these technologies and explore where they could make the biggest impact on patient care.

(60) One potential source of such support could be to encourage more partnerships between hospitals and universities and other research bodies. Efforts are already underway by clinical groups to increase understanding of AI technology – an example is the Faculty of Clinical Informatics, established as the

⁴²⁹You can read more about our Independent Reviewers here: [DeepMind Health's Independent Review Panel, DeepMind, 2017](#)

⁴³⁰You can find all the contracts on our website: [DeepMind Health and Transparency, DeepMind, 2017](#)

⁴³¹[For patients, DeepMind, February 2017](#)

body for UK clinical informaticians to provide resources on the best use of information and information technology.⁴³²

(61) One way to scale this work in medical education would be for medical schools to make clinical informatics part of the medical, nursing and allied health curricula in undergraduate and postgraduate medical training. Appropriate awarding bodies could also consider offering new credentialing and certification in informatics. We are pleased to be supporting Imperial College London, who are leading the NHS Digital Academy, which is equipping IT experts and clinicians with the right knowledge and skills to realise the potential of health IT and AI.

Q8. Are new ethical standards or principles for the use of AI in health needed, or are existing codes of ethics in healthcare sufficient? If new standards or principles are required, what should they consist of? How can we ensure that ethical standards are actually adhered to when designing and using AI for healthcare purposes?

(62) Many of the ethical standards used in medical research are also applicable to AI. Provided that existing robust governance measures for research are put in place for all projects involving AI, we believe existing standards provide a sufficiently robust framework for seeking ethical approvals.

(63) However, AI does raise its own specific ethical issues relating to data handling and proportionality. As noted above, we believe that ethical principles should apply stating that as little data as possible should be used, and that data that is used should be processed in as unidentifiable a form as is feasible for the purposes of the research project under consideration.

(64) In addition, as discussed above, we believe that strong governance, a transparent approach to the use and handling of data, robust audit mechanisms to guarantee the appropriate use of data, and strong legal penalties against attempts to re-identify data, would also be effective at ensuring that ethical standards are adhered to in the practice of AI research.

(65) However, we also believe that further guidance would be welcome from the HRA on the role of AI in the Integrated Research Application System ethical approval process, as well as additional guidance from the HRA's Confidentiality Advisory Group on specific ethical requirements when training models on anonymised datasets.

Q9. What form should Government interventions take, in terms of policy, regulation or investment, in order to help the NHS and society benefit from AI? Should the Government regulate any aspect of the use of AI in healthcare? What in particular, and how?

⁴³²Faculty of Clinical Informatics, 2017

(66) There is a role for government in regulating appropriate access to healthcare data for research, and providing sanctions for those organisations that break the rules. In order to do so, regulatory bodies are likely to need to evolve to deal with algorithmic medicine, as this is such a new field.⁴³³

(67) There is also quite possibly an argument for further investment in capability building, particularly research and PhD funding, in order to ensure that the NHS and society can make the most of the benefits AI could bring through a skilled workforce and the spread of expertise.

(68) We believe that one of the most effective interventions Government could make to help society and the NHS benefit from AI would be to provide funds to allow the NHS to produce open, accessible datasets that Trusts can use to validate the applicability of new algorithms to NHS patient care. By reducing barriers to entry for research, whilst still ensuring safety, security and appropriate controls, the Government could help increase the amount of research and therefore the potential for clinically beneficial AI breakthroughs, which could help patients, clinicians and the NHS.

(69) In addition, we would also recommend fair and transparent guidance be nationally coordinated and offered to NHS organisations like hospital trusts on matters relating to engaging with commercial and research organisations to train AI models.

6 December 2017

⁴³³Digital Health, United States Food & Drugs Administration (FDA), June 2017

Deloitte – Written evidence (AIC0075)

Deloitte submission to House of Lords AI inquiry

Deloitte welcomes the opportunity to contribute to the Committee’s inquiry on artificial intelligence (AI). There are a range of exciting developments in this field that we believe will have wide-ranging impacts on business and society.

AI is increasingly featuring in the conversations Deloitte has with its clients and the services that we offer. We have also conducted extensive research into the impact of technology on the labour market, which we have included a summary of in Appendix A.

We would be happy to assist the Committee throughout the course of this important inquiry.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Advances in computational power and data availability, combined with new techniques and technologies, have driven the rapid increase in artificial technology capability seen in recent years. Nonetheless, it is still early days for artificial intelligence (AI).

Over the next five years, we expect to see significant advances in ‘cognitive automation’. This will result in the automation of repetitive manual tasks that previously required human cognition to perform. Examples of this will be the use of robotic process automation in finance (for example to automate and speed up the processing of information in emails from customers) and the application of AI assisted document reading technologies in sectors such as legal. These are typically high volume, highly repetitive, time-intensive tasks that are either unproductive uses of humans’ time, or that humans do not want to do.

Within ten years, we would expect to see AI become more ubiquitous with ‘always on’ technologies and much more embedded in our day-to-day lives. We have examples of this already, such as voice activated assistants or connected appliances in the home, but technologies like this are likely to become more conversational with users and understand individuals’ contexts. For example, an AI Assistant on your phone will know if you are running late for a meeting, contact the attendees and reschedule if required. Industries such as healthcare will also witness a substantial transformation. For example, AI enabled diabetes health applications will track multiple readouts from a patient, such as blood glucose and exercise, and combine this information with real-time predictive models to help a patient manage their disease. It is likely that cultural attitudes will shift as this advances – for example, there currently remains some

scepticism of medical professionals using technology to assist them but, in future, we are likely to see patients becoming increasingly concerned when technology is **not** being used, rather than when it is.

Twenty years hence, for us, is too far ahead to speculate, but we should expect fully autonomous vehicles, a very high degree of cognitive automation and near ubiquity in the prevalence of Augmented Intelligence – AI helping us in all walks of our professional and personal lives.

Among the factors that could hinder this development are the ability of battery life to keep pace to support 'always on' technologies, data privacy legislation struggling to keep up and a squeeze on the availability of skilled data scientists to drive innovation.

From a broader social perspective, a popular backlash against technological advances, for example preserving certain jobs that could be done by machines in an attempt to delay progress, could also hinder developments in this area. This has been a concern throughout history when new technologies have emerged.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

AI is a fast-moving and exciting field of technology and the level of interest in it, both from businesses and the wider public is understandable. Part of this is around the potential of AI and where the technology may lead us. Certainly, from a business perspective many technology firms are still experimenting with it and have not yet made significant revenues from AI.

On a wider point, as we have seen with technological developments in recent decades, there tends to be significant 'hype' as new technologies emerge and their possible uses are contemplated. However, once they are embedded in our day-to-day lives, this tends to die down and become accepted as normal. We saw this with the emergence of the internet and smartphones, for example, with both now firmly planted in everyday life.

Often the issue is that the wider public underestimate the speed of technological change so developments often appear to be 'fast'.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

The impact of technology on jobs is an area that Deloitte has conducted extensive research into (further details of our series of research are set out in

Appendix A). Our research has led us to conclude that education and skills will be vital to preparing the wider public for the development of technology such as AI and automation.

In our report *Talent for Survival*, for example, we looked at the skills that the workplace, both now and in the future, needs more and less of. Manual skills, raw knowledge and the ability to do repetitive tasks are becoming less in demand, while communication, caring and cognitive skills are becoming increasingly important in the workplace.

The challenge as this pans out is ensuring that we have a curriculum that prepares young people for a world of work where these skills are in demand, and measures in place to enable people already in the workforce to upskill throughout their careers and be able to move across, potentially, multiple jobs in one working life.

At a practical level, we also see the need for more people in the workforce with training in areas such as data science and computing so they are equipped with the knowledge to develop careers in these growth sectors.

On the question of data privacy, from our perspective one of the biggest questions as AI develops is that a subsequent growth, both in number and size, of data sets will pose questions about their security and who owns them.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

This is perhaps still too early to conclude.

From an employment perspective, much has been made of the potential 'hollowing out' of jobs or the creation of an 'hourglass' workforce. However, our research has shown that while jobs are lost to technology, over recent decades there has been more than enough growth in jobs at lower risk of automation and paying more than the job they displaced. The challenge is to ensure people are able to make the transition from 'old' to 'new' jobs.

On a broader point, it should be noted that often the development of technology makes goods and services much cheaper, thereby increasing the number of people able to access them. This is a clear positive for wider society.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Often technological developments are dismissed as 'science fiction' that won't have bearing on our day to day lives, creating issues and fears when, in fact, they do or look likely to. Better educating the public about the role that

technology is likely to play in their lives, how they can prepare for it and where there are opportunities, rather than threats, would help to offset many of the concerns that exist around technology.

From a jobs perspective it is equally important to reassure the public that while technology is changing the world of work - it always has done and will continue to do so - this can be a positive for both jobs and people. Our research has shown that technology removes the mundane, repetitive tasks from jobs, freeing people to focus on much more productive, enriching and rewarding tasks.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

In principle, all sectors of the economy can benefit from technological advances. Certainly developments such as word processing and the emergence of the internet had a significant impact across nearly every sector. On a sector specific level, healthcare is one area in which we see enormous potential for new technologies, both on the clinical side but also in the supporting administrative roles.

The key for any organisation is the ability to fund new technology and experiment with it, often this is a sizeable investment with trialling needed to get it right, which may dissuade firms.

From a business perspective, it is often important within companies to establish who 'owns' technology. Traditionally it would have always been the IT department who were the primary drivers of technology within a business but, in future, we can see technology ownership being much more diverse and set by different teams, such as finance and HR when they look to make use of tools such as robotic processing.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

Our research has focused primarily on the impact of technology on the labour market. Many of the arguments around the 'hollowing out' of the labour market are well known and we have argued throughout our work that education and

skills are integral to ensuring certain groups are not 'left behind' as technology advances.

One factor to consider in the development of AI technology itself is around whether it reflects the biases of those who build it. If, for example, automation is used in human resources or recruitment, does the technology have any biases towards employees or candidates learned from those who designed the system? To be truly fair such systems would be free of these.

There is therefore a critical role for an organisation to step in to provide some ethical guidance, governance and, perhaps, a kitemark system for new technologies to ensure they comply with relevant data protection and privacy regulations. It should ensure that the underlying models and algorithms being used are bias free, representative of society and the data used to train them is appropriate.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

Appropriate control frameworks mechanisms for validation and testing, ongoing monitoring and training of staff all have a role to play. When AI systems generate outputs, it will be important to have an 'audit trail' that can be referred back to see how it arrived at those conclusions and ensure that control frameworks, training, validation, testing and ongoing monitoring processes are transparent and robust.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Government should look at how to foster a conducive broader environment for the development of new technologies, lending policy support as needed to support innovation, skills, access to talent and generally preserve UK leadership.

AI, as a concept, is vast and it would perhaps be impractical for one body to regulate every aspect and use of it. A more productive way would be to regulate it according to its use rather than as a technology in itself. So, when used in medicine or financial services, for example, regulatory oversight could sit with the appropriate industry regulator rather than one all-encompassing technology regulator.

Appendix A

Deloitte has published a number of research pieces looking at the impact of technology on the labour market that we have drawn on in compiling our response to the Committee's questions.

For your reference, these include:

Agiletown: the relentless march of technology and London's response (November 2014) - this forecasted that 35% of UK jobs are at high risk of automation within the next 10 to 20 years, with jobs paying less than £30,000 a year are nearly five times more likely to be replaced by automation than jobs paying over £100,000.

Technology and people: the great job-creating machine (August 2015) – this analysis tracked the impact of technology over 140 years, showing that technological advances have shifted the labour market away from 'muscle power' jobs to care, education and service jobs. Overall the research shows that technology has always helped to create more jobs than it has destroyed.

From brawn to brains: The impact of technology on jobs in the UK (September 2015) – this found that while technology has potentially contributed to the loss of 800,000 lower-skilled jobs between 2001 and 2015, technology has also helped to create 3.5 million higher-skilled jobs in their place. Each of these new jobs was found, on average, to pay nearly £10,000 more per annum than those jobs lost, adding £140 billion to the UK's economy through increased wages.

Transformers: how machines are changing every sector of the UK economy (January 2016) – this looked depth at the potential impact of automation on each sector of the economy and examining how automation and technological advance have impacted job growth and creation within these sectors and which sectors are most at risk, and safe, from automation.

Talent for Survival – Essential Skills For Humans Working In The Machine Age (July 2016) – this analysed the skills and knowledge sets that the UK economy will need in the next fifteen years, showing the increasing importance of cognitive and social skills

5 September 2017

Department of Computer Science University of Bath – Written evidence (AIC0099)

I have already submitted a longer document that I wrote originally for the OECD, titled “Current and Potential Impacts of Artificial Intelligence and Autonomous Systems on Society”, that is fully cited. The below is a brief set of summary answers to your questions.

Please feel free to contact me if I can be of any further service,

Joanna Bryson
University of Bath
Department of Computer Science

The pace of technological change

14 What is the current state of AI and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

AI is now a pervasive technology. For clear thinking about AI policy it is best to take a very simple, straight-forwards definition of AI, as any technological artefact that generates action in response to its own perception of context. With this definition we can see a clear continuous progress from the mechanical governors of the industrial revolution to the “self-learning” systems of the last few years. While machine learning has produced advances that stun us all with their capacity to capture human intelligence, it is important to realise that a) there is a great deal of precedent for what happens each time technology advances our capacity to compute, and b) that computation is a physical process. This latter is important because it excludes one class of alarmist concerns about AI—that one nation, company, or even machine will suddenly create perfect omniscience and thus dominate the world. In fact, laws of computation are laws of nature, and it is provably intractable to know or foresee everything. Computation is not an abstraction like math; computation requires time, energy, and space for storage of intermediate results.

Having said that, AI is already super-human in many domains and in the next 5-20 years it is quite likely that we will be able to capture and express all of extant culturally-communicated human knowledge with it. Already we are far better at predicting individual’s behaviour than individuals are happy to know, and therefore than that companies are happy to publicly reveal. Individuals and parties exploiting this are very likely compromising democracy globally, notably in the UK. There is an incredibly large project here for the social sciences *and* the humanities as we urgently address the political, economic, and existential (in the philosophical sense) challenges of massive improvements in communication, computation, and prediction.

Again, natural laws of biology tell us to anticipate accelerated pace of change given the increased plasticity of increased intelligence. Therefore we need to

ensure our societies are robust to this increase, with sufficient resilience built into the system to allow individuals to have periods out of work finding a new place in the economy. This requires adequate minimum wages, adequate individual savings, and an adequate civil safety net. The greatest decelerators of this process would be 1) war—including cyber/stealth war inducing democracies to dismantle their own critical infrastructures and 2) cybersecurity. The government's present policy of outlawing adequate encryption is a severe threat to the UK on many levels, but particularly with respect to AI.

5. Is the current level of excitement which surrounds artificial intelligence warranted?

See above. Basically, yes, it is if anything belated given that AI is already the core technology of the richest corporations on both sides of the great firewall of China, and given the impact on individual security and on democracy. But no, AI itself is not itself a legal or moral actor and will not take over the world on its own, and there is no particular new threat beyond the damage already done and our increasing reliance on a more-easily-assaulted digital / electric infrastructure. I say again because I cannot understate the importance: having backdoors in our encryption is a substantial security error.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

See first my answer to question 1, which addresses retraining. The most important thing is that we reduce the gini coefficient so that our population retains (or recovers) its social mobility, and those able to innovate have the freedom to do so and the ability to hire others. The productivity and invention intelligent technology should generate should be sufficient to solve the problems of society providing that the economic and political renovations necessary to handle the new redistribution challenges are made.

I am particularly concerned that we are again as in the nineteenth through mid-twentieth centuries in a context of increased inequality and its concomitant political polarisation. We need to remember as we knew in 1945 that it is in the interest of the elite even more than the rest to have a society sufficiently stable to run nations and businesses. The redistribution we practiced from 1945-1978 was not a (successful) war on communism, but rather a necessary economic tactic to counter the technological innovations of petroleum and early ICT. Late (contemporary) ICT requires even greater innovations in shared transnational regulation; the treaties the EU have been experimenting with are not perfect but they need to be improved and extended globally, because the economy is now global.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

It is critical to realise that we have all gained immeasurably from having knowledge at our fingertips. Poor people now have a longer life expectancy than billionaires a century ago. Any talk of “wage stagnation” just tells us how impoverished prices are as an indication of economic value, and how poorly the discipline of economics is serving our society—we need to make massive investment to improve the social sciences. Having said that, and reiterating from question 3, the current aggravation of the essential political problems of a high gini coefficient economy, and also of sustainability, must necessarily be addressed because they threaten stability.

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

The UK is doing an outstanding job of this, a credit to the universities, government, BBC, the Guardian Newspaper, and the Royal Society. We should maintain this level of investment, and probably offer more so particularly through digital university outreach.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Artificial Intelligence affects every aspect of life and all sectors. It is essential that we research how to make AI a standard part of software engineering, and introduce software engineering earlier in education even than A level.

7. How can the data-based monopolies of some large corporations, and the ‘winner-takes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Firstly, although data is very important, I believe that the “winner-take-all” nature of Internet commerce is not just about data, but rather about the relatively low (but by no means zero!) cost of transport of the outcomes of computation. Historically, the cost of travel has been a reliable pressure for wealth distribution—you would not go to the best bakery in the world or even in your county, you would have some individualised function of quality times the cost of travel. New technologies challenge this, whether canals, rail, the exploitation of petroleum rather than coal or wood—each of these innovations required new countermeasures for redistribution.

In addition to this challenge, corporations have learned to evade taxes by bartering in non-denominated ways. Every interaction with Google or Facebook is a barter of information. With no money changing hands there is no tax revenue to support the needs of the global populations facilitating the created value. One of two things has to happen: either we need to find a way to denominate these transactions, or we need to abandon the policy of throttling income with taxation, and turn instead to taxing existing wealth. Although income may be

becoming easier to hide, existing wealth is becoming harder than it has been historically, exactly because of the information age. Economic theory shows that it is far easier to design a stable economy through regulating wealth than through regulating income, but historically this has not been practical because of the power associated with wealth. That this is changing now may be one reason that our democracies have been under such risky assault by the extremely wealthy – perhaps they have good reason to fear that their relative advantage will soon be reduced. However, as I said earlier, creating a stabler society and economy by reducing the gini coefficient and thus ensuring that individuals and corporations cannot destabilise states has the potential to benefit everyone.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

Ethics is the set of behaviours a society uses to maintain itself – as such everything I've said above is relevant to ethics. However, I have above particularly focussed on aspects of safety related to economics and democracy, and only briefly mentioned aspects of safety related to privacy and diversity, so will go into more detail on that here. I will not address consent because I lack expertise in that.

The issues I described in answer to Question 1 concerning prediction are exactly the problems of privacy. It is not only that we do not wish others to know about us, we do not wish others to be able to use that knowledge, and for good reason: because they can then manipulate us. Humanity and human innovation have historically depended on individual diversity which is part of the basis of our notion of dignity. Thus privacy and respect for diversity are both absolutely essential if our society is to prosper, as well as being essential to our individual mental health and wellbeing.

It is important to note that diversity is under assault not only from the misapplication of AI but also from other forms of algorithmisation. I am particularly concerned of the detailed legislation of teaching which reduces the autonomy of individual teachers. This has been generated by a combination of parents' fears of chance events compromising their children's opportunities, and governments' desire to control. In pursuit of equality of opportunity we have generated enforced mediocracy, exactly when what most benefits a citizen is a unique basket of skills, knowledge, and opportunities for insight.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

If you read Frank Pasquale's excellent book "The Black Box Society", the black boxes emerge not so much from AI (the algorithms or source code) as the unregulated gathering and diffusion of data about people. The current system is hopelessly complex in a way we would never permit for money and other legal obligations. There is no question in my mind that AI and ICT more generally have

become sufficiently central to every aspect of our wellbeing that they require dedicated regulatory bodies just as we have for drugs or the environment. However, given that many of these issues have to do with impact on democracy, it is probably not a good idea to have governance only at the national level, since the party in power may well be a beneficiary of any irregularity. Thus I strongly recommend continuing to participate in the EU's world-leading efforts to govern both data, and independently of that, AI.

Please note that saying AI should be subject to regulation and audit is not the same as saying that AI cannot have proprietary IP or must all be open-sourced. Medicine is full of IP, yet it is well regulated.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Yes, please see my answer to Question 9.

Citizens (or perhaps citizens' advocates, see next paragraph) should be able to trigger audits of software systems when they suspect conditions such as a) the inappropriate or unauthorised use of data, or b) unfair or unlawful bias. With respect to data, I advocate for the position that data about a person is a part of a person and belongs to that person. It should be used only for the purpose to which that person has consented. Government regulation and the possibility of audits should encourage companies to use clear, transparent methods to aggregate data and secure methods to store it. With respect to fairness, it should be possible to demonstrate that decisions execute lawful duties and do not disadvantage on the basis of protected characteristics, nor are they arbitrary. Note that the right to audit does not demand that all *code* is transparent, symbolic, or open source. What is at issue is *effects*, so demonstrating valid code is just one possible defense against an audit. Others include: showing that the intelligent system behaves appropriately against a relevant range of inputs, identifying what aspect of an individual's profile produces the contested result, demonstrating a legitimate source of data that results in an output presumed to be based on inappropriately sourced data.

Ultimately it would be ideal if automated systems were in place to answer any individual's complaint or query, but at least initially it will probably be necessary to require citizens to aggregate some threshold number of examples of suspected misconduct before audit procedures are triggered. Again, automated systems might be used to find related filings, and to provide access to already established explanations. Both governments and NGOs should probably be expected to set up such system.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

We should continue participating with the EU efforts. As I mentioned in the introduction, I have also provided under separate cover my 30-page, fully referenced recommendations to the OECD; I hope their final white paper–due out this year–will also be useful.

This evidence is presented on behalf of the Department of Computer Science, University of Bath. It was authored by Joanna Bryson and approved by Eamonn O’Neill (HoD) and James Davenport.

5 September 2017

Department of Computer Science, University of Liverpool – Written evidence (AIC0192)

Response to House of Lords' Select Committee on Artificial Intelligence call for evidence

Katie Atkinson, Danushka Bollegala, Louise Dennis, Frans Oliehoek, Karl Tuyls, Frank Wolter

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Throughout the last 70 years or so, people have been making steady progress, building alternately on insights from neighbouring disciplines such as logic, statistics and optimization. In the last decade, however, there has been a marked change in the typical way that progress has been made, since many large industrial companies have been taking up research in AI and especially machine learning, since it is having a big impact on their companies. As a result of deep learning methods, current AI systems are able to perform many recognition tasks such as image/object recognition at near human-level. There are also breakthroughs made in machine translation, voice recognition and game playing (Alpha Go). Three factors have accelerated this (a) development of algorithms that can modularise the training of deep neural networks (e.g. autoencoders, restricted Boltzman machines, CNNs), (b) availability of large scale datasets for training such models, and (c) GPUs with thousands of computing cores. The datasets available to train such models will continue to grow in the future.

In general, the AI 100 report does a good job in answering this question:
<https://ai100.stanford.edu/2016-report>

2. Is the current level of excitement which surrounds artificial intelligence warranted?

Yes, to a large extent since this excitement is based on scientific achievements. However, caution needs to be exercised on how AI products are reported upon and marketed to ensure that results and expectations are grounded in an evidence base that facilitates further progress and therefore does not hinder take-up through over-hyping.

We also need to distinguish between the excitement in a lot of the academic and tech communities, which is based on an appreciation of the technical strides made in Machine Learning and the application of serious industrial effort to turn many decades of academic research into products, with the kind of excitement (both positive and negative) that we see in the Press and may be communicating itself to the public and politicians. We are still a long way from general, flexible artificial intelligence, but given the recent progress and evidence of achievements in AI, the excitement is warranted and likely to last.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence? *In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.*

There is a key role to be played by knowledge spreading bodies such as universities, schools and the media to inform members of society about developments in AI. We need to provide accurate and scientific knowledge about the methods behind the AI system to the general public so that the public will be “AI ready” in the future. Otherwise, it is inevitable that the public will find AI intimidating and be hostile towards AI systems that would take some of their jobs, while easing the lives of the others. AI researchers should be consulted by governments about policy issues that arise from the successful deployment of AI. Communities of AI researchers are very willing to engage on the societal issues as well as the technical ones; witness the statements in recent years presented at the IJCAI conference signed by leading AI researchers opposing the use of AI-enabled autonomous weapons. Governments also need to be ready to respond to effects of the spread of AI; for example, laws need to be able to handle cases arising from the use of AI systems (e.g. with respect to traffic laws accounting for autonomous vehicles), and also policy decisions, such as how to ensure humans have sufficient means to have a good quality of life if many jobs do become automated (e.g. by considering Universal Basic Income as a means to ensure this).

Many current worries about AI concern “robots that will take our jobs” and this is an understandable concern. The creation of machines that can do many of the jobs we currently do manually equates to the creation of wealth, which is a good thing. How this newly created wealth should be divided is a political question and should be resolved in a manner that is fair to a large segment of our populations. Furthermore with respect to jobs, people often look to the Industrial Revolution as a model. In the long term more jobs were created as a result of industrialisation but we should realise that this is not necessarily a given, though

it seems a plausible hypothesis. However the Industrial Revolution does tell us that changes like this can involve considerable hardship, particularly to those less well-off in society. Re-training and (potentially) re-location of large groups of people are non-trivial tasks with considerable, often detrimental, effects on communities. The government should be putting serious forethought into how such a process could be managed and how the negative effects on the “AI losers” can be mitigated.

The current developments in AI, machine learning and robotics are in principle great developments in capability, but they do entail some possible negative consequences:

Many, if not most, people are ‘data illiterate’: they do not understand the potential implications of the posts they share via social media, as well as other personal data they give away in other forms. Even some of these posts may be very difficult to link back to a person with the current techniques, mining techniques of the future are very likely to reveal identities of former posts, thus enabling companies and other stakeholders to form detailed profiles about large groups of people. This can put those same people at a disadvantage, for instance when taking out insurance, or applying for a job.

Recently, there have been many concerns about ‘fair’ AI and racial and gender biases. This is very understandable, since from a human and legal stance we would want to treat everyone equally (though it should be recognised that this is currently not achieved in many non-AI settings). There needs to be an honest debate of how to weigh the pros and cons of some AI techniques.

With respect to democracy, there are legitimate concerns about the way AI-driven targeting of social media messaging creates “filter bubbles” of opinion and prevents people understanding the perspective of others and accurately assessing the validity and popularity of their own opinions. It is not at all clear how to adequately combat this, though education is an obvious (and probably the least detrimental to democratic process) route. However it is clear already that teaching on “Internet Safety” in schools (which tends towards the alarmist) already lags sufficiently behind many children’s experience of Social Media that it leaves them at best confused and at worst disinclined to listen to advice on the subject from authority figures. We need a more sophisticated approach to such education that understands people make choices between the value to them of access to a service versus control of their personal data, and gives them a deeper understanding of how access to such a service may shape the information they receive.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

A few key companies (Google, Facebook, Tesla etc.) are making the most of the AI developments: they hold most of the data and profit most from this data. To mitigate this disparity between the wealthy tech companies and the general public, we must re-distribute the AI-generated wealth. One means to do so is by imposing an AI tax and distributing it to the public, for example, via Universal Basic Income. In general, a societal and political debate about ownership of data and, (in case of redundancies due to automation) division of wealth, is necessary.

Nonetheless, there is potential for the general public to benefit from the deployment of AI, for example, AI can be used to learn from medical data with this data feeding into decision support tools that have the potential to provide faster and more consistent diagnoses. Similarly in law, AI can be used to support legal decision making in a variety of ways, such as making sense of large data sets and providing support for automated legal argumentation. Another example is healthcare assistants: there are real potential benefits to the elderly in allowing them to remain in their homes longer (which in general has better outcomes). Obviously, initially at least, these benefits are likely to accrue to those who can afford them and those with the energy and education necessary to navigate often byzantine benefit claim procedures in order to gain access to such technologies via the health service. However, long term, widespread deployment of such technologies has the potential to alleviate some of the stress on the NHS caused by the care of elderly patients who, with a little support, could remain in their own homes. Against this there is a real concern that deployment of robots to look after the elderly could increase the isolation suffered by many elderly people.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

The general public is clearly intensely interested in Artificial Intelligence. Lack of engagement is not such an issue, but lack of understanding is. The layman's perceived 'black box' nature of many Machine Learning systems is actually something of a barrier to good understanding - it is easy enough to give a high-level view of how they work, but in the minds of many people, even if they understand the training/optimisation process, the end result still appears to be a magic box into which problems are put at one end and answers come out at the other. This can lead to unrealistic expectations of what such a system can achieve, concern about how reliable it actually is, and the deployment of such systems by the naive in inappropriate ways and applied to inappropriate problems. Clearly part of the solution is more research into questions such as trustworthiness and "scrutability" - part of this may also involve engaging the public with examples of where AI currently fails to give them a better

appreciation of its successes and limitations.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? *In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.*

A few key areas are (though this list is not intended to be exhaustive):

Manufacturing/Robotics is one sector in relation to Industry 4.0.

Chemistry: where AI can be used to assist with the discovery of new materials (a concern of the Materials Innovation Factory at the University of Liverpool).

Healthcare, in a variety of aspects: decision support, robotic assistants (as discussed above).

Law: support for automation of legal tasks and reasoning (see the work of the International Association for AI and Law).

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

This is an important question and it relates to the very important issue that currently the tech companies own the data. If, as soon as one signs up for a new email provider, it would be provided with one's data, this new company could also provide tailored services, thus reducing the "locked in" effect. Another aspect could be to ask companies to adhere to reasonable open standards, e.g., if WhatsApp would be forced to adhere to certain open protocols, it would be possible for other companies to develop their clients that people could use without disconnecting from their networks of friends.

Another option is that the data of a user must be owned by that user and stored in a cloud storage dedicated for that user. A company may use that data to provide a service that the user registers for and provide explicit consent for this.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? *In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.*

The answer to 3 above is of relevance here. Within this topic it also needs to be considered how reliable human decision making is compared with decision

making in AI systems. It may well be that humans end up performing worse in some decision making tasks than an AI system due to a variety of reasons such as humans' inability to assimilate datasets as large as AI systems can, humans' unconscious biases and all sorts of physiological factors such as tiredness that reduces the ability to make good decisions. Of course, AI systems also need to be tested rigorously to ensure that biases from training data are recognised and addressed. And we should be clear about domains where human judgment is critical (e.g. weapons deployment) and those where it is less so (e.g. routine office tasks).

AI-based systems that are integrated into our everyday lives will need to embody and reflect our values - (driverless cars will need to reflect societal expectations of polite behaviour as well as obeying the Highway Code, robots in the home will need to respect the normal of behaviour, privacy, dignity and comfort of the occupants, and so on). There is also a clear question about who gets to decide what values are embodied in this system, and how this is to be achieved, and there is a concern that ad hoc approaches developed on a company by company basis may not actually reflect society's expectations and could even allow the technologies we build to support entirely unethical groups.

On issues of privacy and consent, one problem is allowing people to have an accurate appreciation of the ramifications of any consent they give. This is not an issue exclusive to AI, but if we have AI systems communicating on our behalf with other systems, then the information they share and what is done with it becomes a part of the problem.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

The term "black box" suggests a binary classification has been used to abstract away from the simple truth that capturing real life relations is complicated, because these relations are complicated. Neural networks (or any Machine Learning algorithm for that matter) are not black boxes in the sense that you can inspect every single neuron in a neural network if you wish to do so. Humans have a built in mechanism that makes them comfortable to treat anything they cannot understand as a "black box". "Black boxing" is a way of abstracting concepts and treating them as monolithic units to build even more complex systems. Rather than trying to ignore this by banning 'black box' models, it would be more productive to focus on (developing methods that help) understanding the complexities. More generally, it is recognised currently that transparency is an important issue in the ongoing development of AI systems and this is something that researchers are keen to address. On developing methods to assist with the explainability of AI systems, a variety of techniques are being developed, such as visualisations to communicate results and

computational models of argument to justify decisions.

Referring back to the ethics question above, where we want a system to weigh the competing claims of say efficiency, comfort, privacy and safety in some complex situation, we need some mechanism which will allow us to understand its choice in terms of the varying priorities of those values. This could well be achieved through better integration of machine learning with other systems (e.g., argumentation or logic based reasoning) in order to provide transparency at an appropriate level.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Regulating AI at this stage is likely to obstruct its development. Rather, the government could actively engage in AI development by funding universities and companies who are working in AI, thereby “investing” in this field. That way the government will have better control in the long run in this field, and could regulate if it wishes so.

However, the government will need to give consideration to appropriate legislation, particularly to decide upon who is responsible for the behaviour of an AI system (especially any that may perform online learning based on the behaviour of users), what constitutes an appropriate level of care of users from the developers of commercial complex systems, and the standards of behaviour to which such systems will be held.

At the same time the government should be giving serious thought to the potential effects of AI, particularly on already disadvantaged communities, of major upheavals in the job market, and how those effects can be managed so that both the benefits and downsides can be as equitably distributed as possible.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

The US is a good example for learning lessons on AI. In the US, it is major companies who are the major funders investing in AI. This level of industrial commitment does not happen in UK (even in the EU or Japan). When a company is investing in an AI project, there is a clearer focus than with a government funding. The stakeholders and accountability constraints are stricter with

company funding, which forces more tangible outcomes. It also enables the projects to out-live the funding period and many researchers to share common visions/goals (potentially set by the company who employs them).

The IEEE is currently running a Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems which is considering a wide range of issues from an international perspective and is spinning out a number of standards working groups on, for instance, Transparency, AI guardians, Ethical Design processes and the like. Germany has just released ethical guidelines for self-driving cars. ISO is also working on standards for AI “Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)” and “Artificial Intelligence Concepts and Terminology”. The BSI has already led the way with its standard for Robots and robotic devices to guide the ethical design and application of robots and robotic systems. It probably isn’t necessary for the government to seek to develop independent UK standards for AI, but it should be tracking and participating in the development of these international standards and forming a view on how they might inform certification processes for robotic (and other AI) systems in the UK.

6 September 2017

Digital Catapult – Written evidence (AIC0175)

1.0 INTRODUCTION

The UK is viewed as a global thought leader and centre of skills in Artificial Intelligence and Machine Learning (AI/ML). The country has 4 of the top 20 global universities for Computer Science, along with being 2nd in the world after the US for influence of AI Academic Publications, 3rd globally by number of publications in AI/ML (behind US and China) and 4th globally by number of citations in AI/ML (behind China, US and India) according to the [Schimago Journal and Country Rank](#).

When this is combined with a concerted investment in R&D activities in this area, with over 144 EPSRC grant funded projects across 47 institutions, alongside the setting up of the Turing Institute as a world-class research centre, the Leverhulme Institute for the Future of Intelligence and AISB: Society for the Study of Artificial Intelligence and Simulation of Behaviour that was founded in 1964 – it is clear the UK has a rich history and an exciting future in this space. Indeed, by December 2016 there were over 50 AI/ML and Data Science meetup groups and communities across the UK, with over 41,000 members in London alone.

This base of talent has led to a growing AI/ML SME ecosystem, an active VC community and some of the Europe's most high-profile acquisitions taking place over the past few years. This includes: Google's purchase of DeepMind in 2014 (\$500M); Microsoft's purchase of Swiftkey in 2016 (\$250M); Twitter's purchase of Magic Pony in 2016 (\$150M); and Apple's purchase of VocalIQ in 2015 (\$50M). Furthermore, [based on recent Crunchbase data](#) total funding for the top 10 AI/ML companies in the UK such as Improbable (\$500M), Dark Trace (\$179.5M) and Benevolent AI (\$100M) equates to over \$1BN.

The UK's AI/ML startup and SME ecosystem leads by fair margin across the EU too. Through research conducted here at Digital Catapult we have identified over 600 AI/ML companies in the UK, which equates to almost half of the 1,249 AI/ML companies across Europe. On the global stage, although the UK is behind Silicon Valley, which has nearly double the number of startups and SMEs in this space, the EU as a whole is also ahead of Asia. However, with the significant investment in Deep Learning research by China as demonstrated by their recent commitment to making China a world leader by 2030 and having major breakthroughs [by 2025 building a 1 trillion yuan AI industry \(\\$147.9 billion\)](#) – it is a matter of time before Asia catches up on the commercial front and as such the UK must ensure it builds on its existing advantages to keep pace.

2.0 WHAT IS ARTIFICIAL INTELLIGENCE?

AI was coined as a term at the 1956 Dartmouth Conferences by which John McCarthy and a number of other influential computer scientists. However, AI achievements were slower than the ambitious expectations at the time, which led to an “AI winter” - a significant slow-down of investment, research and interest in the field. In recent years, advances in the field of machine learning have led to the current AI renaissance, with the technology being championed as the key to maximising global productivity and solving the world’s challenges, through to being dismissed as an asinine idea or the biggest existential risk facing humanity in modern times.

The reality is that in the past few years (particularly since 2015) despite the wide recognition of both its opportunities and risks, AI has grown exponentially. This is widely due to the increased availability of faster, cheaper and more powerful computation, alongside a flood of data and affordable / scalable data storage to facilitate machine learning models development and inference. At the same time, it is worth considering the depth and breadth of UK expertise of additional AI sub-disciplines (including symbolic AI, simulation, computer vision, NLP, machine learning, robotics etc.). The continued success of AI will require all of these to thrive (as opposed to just end-to-end deep learning).

Most people when discussing AI think of General AI - machines that are able to replicate human reasoning, thought processes and analysis so that it possesses the same characteristics as human intelligence across a full range of cognitive tasks. In reality, we are still decades away from “C-3PO” or “Bladerunner” levels of artificial intelligence. However, where we are seeing significant progress, adoption and excitement is in the form of Narrow AI - where technologies are able to perform and even out-perform humans in specific tasks. The most famous example of this is its use in strategic games – with programs such as AlphaGo, Google DeepMind’s narrow AI program built in 2015 that last year beat the world number one Go player.

Narrow AI is utilised in a diverse set of applications such as language translation, self-driving vehicles and image recognition. It has become the backbone for the commercialisation of AI with companies such as Google, Netflix and Amazon using it to provide recommendation systems and advert targeting. Important progress is being made in its use-cases in healthcare in the form of diagnosis and medical research. In manufacturing, it is being used in predictive maintenance of machines and optimisation of the supply chain. In the creative industries, companies are using it to build immersive worlds in virtual reality, and even to compose creative content of its own. It is from this increased exposure that AI is capturing the public imagination – with leading global figures such as Elon Musk and Stephen Hawking in recent years raising concerns and issuing public warnings about the development of general artificial intelligence. However, although we agree that long term consequences should be taken into

consideration, Digital Catapult takes a pragmatic view on Artificial Intelligence, its potential to transform the UK economy, increase productivity and drive new opportunities for growth will be considerable in the next 3-5 years. Solutions to some of the deeper ethical concerns may not be so critical in that time frame.

3.0 WHAT FACTORS WILL IMPACT THE DEVELOPMENT OF AI/ML

Digital Catapult believes that the UK Government should consider the below immediate challenges in developing and deploying AI/ML based products and services:

- *Access to Data, international data flows, data markets and the danger of data monopolies.*
- *Access to Computation Power.*
- *Adoption of AI/ML.*

3.1 ACCESS TO DATA AND DATA MONOPOLIES

To create new products or services that employ machine learning techniques, organisations need numerous examples (“training data”) for the algorithms to learn from. Indeed, it is the availability of large training datasets that has been the bottleneck and one of the most fundamental challenges to the growth and adoption of AI/ML technologies across the economy.

Significant progress is being made on this front by global technology companies who have unparalleled access to training data. For example, Google Translate achieved a breakthrough performance at Arabic and Chinese-to-English translation using a dataset with more than 1.8tr tokens from Google web and news pages. It also has a huge pipeline of incoming data, currently translating over a 100bn words a day. Furthermore, Facebook’s deep learning face recognition system was trained on the, “largest facial dataset to-date, an identity labelled dataset of four million facial images belonging to more than 4,000 identities”.

Furthermore, companies are beginning to open source their algorithms and concentrate on effective forms of data collection. Examples include TensorFlow from Google and Torch, contributed to by Facebook, as well as Facebook’s Big Sur hardware designs, while Amazon contributed to the \$1bn invested to found Elon Musk and Sam Altman’s new non-profit research company OpenAI. As machine learning algorithms become commoditised we are also beginning to see the development of ‘AI as a Service’, with examples such as Microsoft Azure, IBM Watson, Google Cloud Machine Learning or Vision. For each of these organisations, there is more value in the data than the algorithms, thus their access to such large data sets gives them significant advantages over smaller companies.

Despite the beneficial access to these datasets by larger companies, it is often forgotten that it is also critically important for growing startups to gain access to similar growing pools of training data in order to maintain economic growth, disruption and further innovation. For AI/ML SMEs it is vital to the development of their products offering and is a key defensible asset. Indeed, often smaller companies are evaluating merger or acquisition opportunities based on the availability of data and are only interested in an acquisition if the acquirer has the right dataset.

Acquiring sufficient training data is therefore a significant challenge for AI/ML SMEs and as such we are seeing the emergence of data access strategies and new business models built out of the necessity of relevant and useful data. Moritz Mueller-Freitag has categorised these strategic interventions into 10 varieties, ranging from creating datasets by hand to releasing 'side' applications that are valuable to consumers but have a side effect of generating large sets of training data. However, while many AI/ML innovators are looking to build a "data network effect" larger companies who have access to this data are already at a significant advantage. A normal network effect will make a service more valuable as it acquires more users. The more people use a social network, the more valuable it is to each user. This tends to lead to a 'winner takes all' market. A data network effect happens when a service becomes smarter (through acquiring more data) the more people use it:

"the more users use your product, the more data they contribute; the more data they contribute, the smarter your product becomes ...; the smarter your product is, the better it serves your users and the more likely they are to come back often and contribute more data – and so on and so forth. Over time, your business becomes deeply and increasingly entrenched, as nobody can serve users as well."

– Matt Turck, Venture Capitalist at FirstMark.

This effect will mean that in many markets global technology companies already have an inbuilt set of advantages: they have existing consumer and business relationships, they have large messaging and social networks, they have internet content, they control the large mobile operating systems and see usage data across many platforms, products and services. These so called "data monopolies" are using their dominant data position to attack new markets such as the internet of things, autonomous vehicles or healthcare (for example, Google has Nest, Sidewalk Labs, the Self-Driving Car Project, DeepMind Health; Apple has HomeKit, HealthKit, CareKit, CarPlay; 5% of US Amazon customers now have an Amazon Echo device listening for voice commands in their homes). Larger players lacking in data will make acquisitions to catch up, such as IBM's recent acquisition of the Weather Channel and Truven Health Analytics. Digital Catapult has written about the data network effect for smart cities, where we

believe that city data and resulting services will also be dominated by GAFAs companies, rather than local governments or Internet of Things vendors. We are also due to publish an in-depth report on the importance post Brexit of the UK positioning itself in relation to data market opportunities to encourage the growth of data markets and not inhibit them, due to be published imminently.

Looking at the 10 specific strategies to acquire data for machine learning as suggested by Mueller-Freitag, there are really only three where a startup is not at a huge disadvantage:

- Narrow domains that have not yet been addressed
- Using publicly available or licensing third party datasets, although these will form less of a defensible asset
- Collaborating with large corporations to solve their problems and access their data. Here there will be cases where startups may have an advantage as the big internet players may not be trusted by large corporations. However, in other cases precisely the opposite will happen (for instance the recent arrangement between the NHS and Google Deepmind).

We would add a further factor to that is not listed in Mueller-Freitag's 10 strategies. Despite the global nature of internet markets, it is not currently the case that all data can flow freely across borders. Personal health data, as in the example above, can be tightly regulated and restricted to data centres in specific countries; defence or security data even more so. Although this is a potential advantage for local startups or scaleups, in practice the NHS/Deepmind arrangement and Google's local presence in the UK provides a counter example.

We can see that in the majority of cases the ready access to funding, distribution channels and existing data sources gives the incumbents huge advantages, and makes it extremely difficult for startups to scale. Furthermore, the data network effect is even stronger than we've described, as it also serves to pull in more of the machine learning specialists who are in such short supply. They are attracted to the organisations that can provide them the biggest datasets and largest user populations.

The level of data collection and ubiquity of these large tech companies have led to calls and debates in recent months for tech giants to be broken up in order to create a more level playing field. While these debates are on-going, just this week we saw a partnership being formed between Amazon and Microsoft around their AI voice assistants Alexa and Cortana. Although one could argue these tech giants are competing with each other for market share, partnerships such as this make their reach even broader and potentially stifles the market for innovative SMEs in this space even further. However, the EU recently delivered a \$2.7 billion fine to Google in late June against Google for favouring its shopping comparison service over rivals in search results. Furthermore, when coupled with

Japan & South Korea's questions around both Facebook and Google's ability to collect web-surfing and online purchasing data from more than a billion people, we are beginning to see regulatory / public policy reactions as part of a growing concern against these actions.

There are arguments for and against the break-up of these large internet companies. They don't engage in typical monopoly behaviour such as stealing market share by selling goods below the cost of production and their success has arguably benefited consumers (few people would be happy to give up using Amazon's one day delivery or Google search) – not to mention the majority of their products and services are free. However, data is becoming even more valuable and sought after than ever before because of AI/ML and proactive regulatory changes need to be made to ensure that these companies do not restrict SME access to data and the AI/ML industry is able to expand significantly as a result.

3.1.3 Data Access Recommendations

When it comes to commercial entities and access to data, Digital Catapult recommends:

- 1.** Having measures in place to encourage companies to collect and store their data in a usable way, and making it available for analysis.
- 2.** When public organisations (such as healthcare providers) negotiate IT deals, we recommend that in addition to maximising the utility of the immediate tech solution, data related issues should also be taken into account: ownership and access to current as well as future data.
- 3.** Value of data in such deals is frequently underestimated. Possible solutions include standardising measures of quantifying that value, as well as data value related training for procurement professionals.

3.2 ACCESS TO COMPUTATION POWER

3.2.1 Cost of Expertise and Infrastructure

Even if one assumes that the cost-element of access to computational resource can be solved, significant expertise is still required to build the infrastructure for machine intelligence research and deployment pipelines.

“You need to build systems to run very large, demanding jobs at scale, and to do this in an easy-to-use way so your researchers can conduct as many experiments as they desire. These parts are not commoditized and when you move into AI systems that require larger and larger models, the expertise required to make the infrastructure grows. It doesn't diminish.”

Jack Clark, [The Public Policy Implications of Artificial Intelligence](#),
OpenAI

Acquiring this expertise is both costly and time-consuming.

3.2.2 Recommendations & Digital Catapult Activities on Access to Computation Power

Digital Catapult as a result are building a Computation Lab that will enable SMEs to gain access to computation power, by either:

- 1.** Increase capacity through subsidising computation cost for SMEs
- 2.** Providing access to national high-performance computing infrastructure.
- 3.** Encouraging collaboration with UK SMEs producing new hardware infrastructures.
- 4.** Investing in research into modelling methods that require less computation power (and/or data).
- 5.** Developing hardware and system expertise; and supporting benchmarking efforts

3.3 WIDER ADOPTION OF ARTIFICIAL INTELLIGENCE & MACHINE LEARNING TECHNOLOGIES

Adoption of existing AI/ML solutions by larger organisations is still slow and sometimes lagging behind other countries. We would recommend to add programmes for education around possible AI/ML solutions as well as encourage sourcing these solutions from SMEs. Through the establishing of a comprehensive national programme around Artificial Intelligence, there can be more structured engagement, encourage greater collaboration across academia, large companies, SMEs and government to bring about a better understanding of the opportunities across industries.

To this end Digital Catapult fully supports the identification of AI/ML as a game changing technology in a number of the ongoing Industrial Strategy Sector Deals and ISCF bids, including the Industrial Digitalisation Review that Digital Catapult have contributed to as a member of the Leadership Team and have helped to design the national ecosystem that underpins the delivery mechanisms to inspire greater adoption of AI/ML and further innovations in an applied challenge focused setting. We also fully support the recommendations around AI/ML within the context of the Life Sciences Review, and the more in-depth reviews ongoing around AI/ML in the UK (led by Digital Catapult’s non-executive Director Wendy Hall and the CEO of Benevolent AI, Jerome Pesenti), and the Robotics and Autonomous Systems review.

To encourage adoption of AI/ML technologies Digital Catapult is a strong proponent of Open Innovation events, to bring together companies from across

the economy into a collaborative environment with innovative start-ups, scale-ups and SMEs who are developing AI/ML commercial solutions, along with forward thinking academics and researchers. This bridges the cultural divide between larger companies and innovators, encouraging greater adoption of the technology and encouraging them to create more fertile environments to work with together such as sharing labelled and useful data within a manufacturing setting.

4.0 WHERE CAN AI & ML POSITIVELY IMPACT THE ECONOMY AND SOCIETY?

Digital Catapult recognises both Digital Manufacturing and Health as two major markets to be disrupted by AI/ML - but we see it is also as having a substantial impact on the economy across the board including applications in creative content generation for the creative industries which is demonstrated by the recent \$500m funding for Improbable in its use of AI generated immersive content.

4.1 DIGITAL MANUFACTURING

Digital Catapult see strong potential in the use of AI/ML in Digital Manufacturing and it is for this reason that we are part of the leadership team and project management office for the Industrial Digitalisation Review Sector Deal being led by Juergen Maier (CEO of Siemens UK) in shaping the adoption of emerging technologies by Industry. Its uses include the below:

- 1.** By utilising data collected from internet of things sensors and equipment across the factory floor and the supply chain, machine learning can be hugely beneficial to manufacturers in the form of Predictive maintenance, production optimisation and streamlining of processes.
- 2.** AI and Machine Learning can support the growth of a more decentralised, localised and personalised strand of manufacturing, that utilises data and autonomous machines to directly connect consumers with makers. This could build new business models and opportunities to transform the economy.
- 3.** By opening up data sets to machine learning SMEs and researchers, manufacturers across the supply chain will be able to utilise the technology to transform their business and increase productivity significantly.
- 4.** We believe the UK - with its strong background in AI/ML research - is well positioned to become a global leader in Industrial Digitalisation, positioning itself as the go to place for the underpinning technologies required for industrial digital transformations that could lead to substantial economic benefits.

4.2 HEALTHCARE

There are significant opportunities for the UK in healthcare applications, including opportunities for personalised medicine, harnessing data to get a deeper understanding of the causes of disease.

- 1.** Machine learning has the potential to transform healthcare, enabling better medical decision making as well as novel methods for prevention, diagnosis and treatment, while at the same time creating new global opportunities for UK machine learning companies
- 2.** The UK's health system will improve both health outcomes and productivity by applying machine learning technologies, but in this fast-moving field this has to be in partnership with smaller innovative companies or research groups. This means opening up the flow of data between the health system and the innovators, and in particular opening up access to training data for machine learning
- 3.** We have seen examples of this in action recently with DeepMind and two London hospitals. This is encouraging and we would like build on these examples to enable competitive advantage to UK innovative SMEs
- 4.** We believe the UK is well placed to provide leading innovation in the areas of personal data, trust, security, privacy and the ethics of decision making by AI algorithms, that would help address the barriers to our future personalised and decentralised health system

DIGITAL CATAPULT'S ACTIVITIES IN ARTIFICIAL INTELLIGENCE/MACHINE LEARNING

- 1.** Building a world leading Computation Lab to reduce overheads and initial costs for high potential AI/ML SMEs in the UK.
- 2.** Establishing a programme of data challenge competitions for UK based AI/ML SMEs to gain access to data sets from across the Digital Manufacturing, Health and Creative Industries.
- 3.** Bringing together AI/ML Researchers with entrepreneurs and experts in the Manufacturing, Health and Creative Industries to help them to fully and realistically understand the potential and value of the technology.
- 4.** Building a comprehensive ecosystem map of the AI/ML landscape in the UK.
- 5.** Positioning the UK as a world leader in AI/ML through white papers, global missions and dissemination of UK ecosystem best practices.

6 September 2017

Doteveryone – Written evidence (AIC0148)

Written evidence submitted by Doteveryone, a think tank fighting for a fairer internet, 6 September 2017.

1. Definition of AI
 - 1.1. Doteveryone considers AI to encompass machine learning, algorithms and neural networks – driving changes in tasks and occupations, business processes, business models, power structures, and wealth (including informational and infrastructural as well as monetary.)
2. Public understanding and AI (Q3, Q5)
 - 2.1. The future of AI is subject to a lot of speculation about potential and likelihood, with future scenarios for AI ranging from the fantastical to the cataclysmic.
 - 2.2. The public needs more realistic, nuanced and balanced communication about what is likely to happen in the field and application of AI today, tomorrow, and 10 years hence.
 - 2.3. The starting point for public contribution to the wider AI conversation is to ground it in relatable, familiar examples of where AI is already part of people's lives. For instance, in insurance premiums, price comparisons, or cancer diagnosis. Then discussion and debate can move forwards to future applications and implications.
 - 2.4. The best preparation the general public can have for AI, and indeed any technological change, is to have digital understanding.
 - 2.5. Doteveryone defines this as the ability both to use technology and to comprehend, in real terms, the impact that it has on our lives.
 - 2.6. Where digital skills enable people to use digital technologies to perform tasks, digital understanding enables them to appreciate the wider context of and around those actions.
 - 2.7. In the case of AI, understanding would mean the public being aware of which systems, processes, decisions, tasks, and interactions in their lives involve AI, and what influence it could have into the future.
 - 2.8. The public should be helped in gaining this understanding by the organisations that use, and might use, AI. They should be required to declare where and how they use it, similar to declarations of the use of CCTV or telephone recording.
 - 2.9. Organisations should also be open and responsive to questions about their AI use, and ready to explain its functions and outcomes in plain terms.
 - 2.10. There is also an important role for active and empowered civil society, whose resources and expertise the public needs for support and representation. Specifically, Government should allow consumer groups and activist bodies to bring class actions against AIs.

- 2.11. Doteveryone has begun creating a definition of digital understanding, to decide what its fundamental elements are, so we can find ways to help the UK's internet users build them.⁴³⁴
- 2.12. Doteveryone has also been exploring how these and other aspects of responsibility and accountability could be built into technology, including AI,⁴³⁵ and how a consumer mark for trustworthy tech could operate, as an aid public understanding.⁴³⁶
3. Leadership and AI (Q3, Q5)
 - 3.1. The public can only be properly supported in their understanding if policy-makers, legislators and civil servants have the digital understanding necessary to robustly debate and scrutinise AI issues.
 - 3.2. It is important that leadership on AI is shared throughout any organisation deploying AI, so it isn't the sole concern of a siloed, dedicated individual or team. This will increase the diversity of input into the use of AI, the appreciation of its possible impact across an organisation, and the potential for spotting problems.
 - 3.3. Doteveryone has begun a digital leadership programme to show key decision-makers how they can develop the digital understanding they need to be better leaders in the digital age.⁴³⁷
4. Public versus private value from AI (Q4, Q7)
 - 4.1. The current beneficiaries of AI development and use are the organisations with access to the most expertise, data, and computer hardware – mostly private companies.
 - 4.2. We are seeing the benefits of using AI to solve problems in the private sector. The public sector, too, should be harnessing this potential where appropriate, in areas which involve public consultation.
 - 4.3. Government needs to access its own AI expertise, and develop its talent and capacity, rather than waiting for or relying on corporations to unlock potential.
 - 4.4. This will involve the development of technological infrastructure, in a responsible way, and, again, the building of strong digital leadership.
 - 4.5. Outsourcing Government AI and data work to corporations, with their ready infrastructure and expertise, is tempting for its speed, ease and low upfront cost. But doing so runs the risk of putting potential public benefits into private hands and tying Government into using expensive AI systems with potentially rising costs.
 - 4.6. There is huge potential for improved efficiency in Government and the public sector if AI is used effectively, which would lead to huge savings in public money.

⁴³⁴ Doteveryone, [This is digital understanding](#), 05.09.17

⁴³⁵ Doteveryone, [What is responsible technology anyway?](#), 13.04.17

⁴³⁶ Doteveryone, [A trustworthy tech mark](#), 31.08.17

⁴³⁷ Doteveryone, [Helping leaders understand digital](#), 20.04.17

- 4.7. The near-term costs of doing this are not insignificant but the long-term economic benefits of building the UK's AI capability, for the shared benefit of the population, would make the investment worthwhile.
 - 4.8. Government holds and collects a large quantity of good-quality data. Public data that has public value, but isn't openly shared, offers even more value when processed with AI techniques. For instance, NHS data used appropriately and with appropriate patient involvement, could develop and advance healthcare.
 - 4.9. Access to this data must be granted in ways that ensures public benefit – both in outcomes and, where appropriate, financially – from its processing and from the new tools and products developed as a result.
 - 4.10. Government should also be ready to explore, and open-minded in exploring, new organisational models for developing public AI, such as cooperatives or government-owned entities. Silicon Valley corporations and UK startups are not the only options.
 - 4.11. AI, like other technologies, is a tool – it offers power to those who hold it, through information and understanding. Public debate about it should not only be about good versus bad, but about who controls that power, who the power is over, and about what structures for accountability and redress do and should exist.
5. Sectors and AI (Q6)
- 5.1. The tech industry has a history of solving the problems of young, affluent, white men, rather than tackling wider social issues. For example, there are hundreds of apps and wearables for tracking fitness, but very few for tracking menstruation;⁴³⁸ some automatic soap dispensers will only work for light skin.⁴³⁹
 - 5.2. Lack of diversity among technology developers and business leaders leads to a lack of diversity in the problems they solve, the ways they solve them, and the solutions they find. AI is no different.
 - 5.3. If the data for AI isn't available for a sector, or the data that is available doesn't capture its breadth, there's a risk that we focus on only what can be measured, and build from there – we overlook the value in things that are hard to measure or can't be measured.
 - 5.4. Important sectors, like care or education, are unlikely to benefit from AI in the near term, or are more likely to be subject to inappropriate AI use, because of AI's dependence on data.
 - 5.5. There is also potential for bias in AI decision-making from biased input data, or from human interaction with algorithms.
 - 5.6. Much historical data either describes a past where biases were prevalent that we find unacceptable today, or it was recorded using methods and means which reflect the attitudes and perceptions of the time.

⁴³⁸ The Atlantic, [How self-tracking apps exclude women](#), 15.12.14

⁴³⁹ Evening Standard, [Automatic soap dispenser sparks 'racism' outrage](#), 18.08.17

- 5.7. An example of these elements potentially affecting algorithmic decisions comes from the US, where research indicates inaccurate, past assumptions about neighbourhood demographics may be feeding into US drivers of similar risk paying different premiums.⁴⁴⁰
- 5.8. It is important to remember that AI is a tool, not a silver bullet. We should use it to help us solve problems only if appropriate – namely where needs have been identified but traditional approaches to meeting them haven't worked.
- 5.9. We should also treat AI as we do other tools in that it can be overruled if there is an exceptional case, where there is a right to appeal and redress.

6. Responsible AI (Q8)

- 6.1. Very little legislation has been agreed around the ethics of AI. Most discussion happens between experts rather than in open arenas.
- 6.2. As a democratic society, we need our Parliamentarians to have open discussions that include and inform the public, as AI policy and legislation for the UK are built and shaped.
- 6.3. We need transparency in AI decision-making, so both the human and machine elements can be identified, explained and understood.
- 6.4. There also needs to be full consideration of how organisations, businesses and others will be held accountable for the power they wield through AI, and of how liabilities are assigned across those accountabilities.
- 6.5. A code of ethics and professional standards should be required of those developing and deploying AI, and indeed other advanced digital technologies. This should be combined with a tangible, statutory requirement for transparency and accountability which acknowledges the human involvement at AI's foundations. 'The algorithm did it' is not an acceptable response.
- 6.6. These codes, standards and requirements should apply to both private and public sectors.
- 6.7. There are legal considerations around public data collected by Government or its agencies, but not open data: namely who owns it, and what can legally be done with it, especially when publicly collected data is also personal data.
- 6.8. In-house Government AI work will also require a change in attitude and approach to the sharing of all data across separate, and often separated, areas of Government. This should happen in a way that responds to and satisfies public concerns about data use.
- 6.9. The significant breach of data law between the NHS's Royal Free Hospital and Google DeepMind exemplifies many of the major issues at

⁴⁴⁰ ProPublica, [Minority Neighborhoods Pay Higher Car Insurance Premiums Than White Areas With the Same Risk](#), 05.04.17

stake: the lack of competence public bodies have in negotiating AI agreements with the private sector; the potential for harm to privacy rights and public trust in data transfers; and the giving away of valuable public data assets to private companies for free.⁴⁴¹

6.10. The implications around ethics, standards, liability and accountability are and will remain an evolving issue – we won't know the full implications of some of these decisions into the future. Legislation needs to be framed so issues can be dealt with as they arise.

7. The role of Government in AI (Q10)

7.1. Public awareness of AI and understanding of its impact must increase. The Government has a huge role to play in resourcing and supporting efforts to do this.

7.2. Government must do more to balance national needs with the demands and expectations of global entities involved in the development of AI. It must seek conversations with experts outside those global organisations, engage with the public and representative groups, and be transparent about input from and lobbying by technology businesses.

7.3. The ICO, and other ombudsmen and regulatory bodies that exist or may arise, will need to be properly resourced, empowered and incentivised if they are to fulfil their roles in a digital society. Their remits are only likely to grow over time as AI and its use of data expands.

7.4. For instance, in the case of GDPR and the Data Protection Bill, an adequately resourced ICO means it being given clear incentives to act, to underscore its powers.

7.5. Cyber safety and security around information must be on a par with other AI and data concerns. It is a cross-border issue – there is no firewall around the UK.

7.6. Government must continue to coordinate internationally, and, particularly, remain engaged with European data, safety, and computer and network security activities through ENISA and other agencies.

6 September 2017

⁴⁴¹ ICO, [Royal Free - Google DeepMind trial failed to comply with data protection law](#), 03.07.17

Reverend Dr Lyndon Drake – Written evidence (AIC0108)

Submission made in an individual, personal capacity.

1. I suggest that making good ethical judgements about the use and impact of AI systems in society often requires significant knowledge of the field, particularly around questions of black-boxing and data privacy.
2. As background, I have a PhD and a number of peer-reviewed scholarly publications in the field of AI, although it is some years since I have been actively involved in research. I previously worked for an investment bank, and during that time I wrote a rudimentary automated trading system. Since leaving academic computer science I have continued to follow developments and some scholarly literature.
3. Many of the issues around AI are well-known. Nevertheless, commentary on the impact of AI systems and the ethical issues which arise from the increased capabilities of these systems sometimes lacks technical precision.
4. For example, while AI systems have become rapidly more capable of carrying out tasks that until recently required human intelligence, the ethical questions raised by AI systems can depend to a considerable degree on the AI methods used.
5. Most of the recent advances owe their progress to the availability of extremely large amounts of data, generally tagged by humans, and often obtained as a side-effect of offering an attractive service at no up-front cost to users. Facebook and Google are obvious examples. (By contrast, Apple makes a marketing virtue of the privacy of its services, which it charges for, and its AI services are generally considered to be less effective than those of competitors). These systems generally use some form of Machine Learning (ML), often Artificial Neural Networks (ANNs). Many other AI techniques exist, but ML approaches (and ANNs in particular) have become dominant because of the availability of very large training datasets (and to a lesser extent, because of the cheap and ready availability of computing resources through cloud platforms).
6. ML approaches generate rules from the datasets used for training. The rules they generate are typically impossible for humans to interpret or modify. Given sufficient amounts of training data of sufficient quality (including accurate tagging of the training data by humans), ML approaches can be highly successful within a particular domain. The tasks of computer face recognition or automated driving can be carried out very accurately by systems that make use of these techniques. Unfortunately, these approaches also have significant drawbacks which must be considered when weighing up the ethical impact of AI on society.
7. The datasets required to train many of these systems are so large that the only practical way to obtain sufficient quantities of training data is from humans voluntarily making personal data available. It is not obvious that the

users of Facebook and other services are really aware of the extent of the use of their personal data, and in particular the way their online actions and content are being encoded in AI systems. Admittedly, this encoding is carried out in ways that make it extremely unlikely that any of that personal information could ever be extracted back out of the AI system. But that does not remove the necessity of real consent by those whose personal information is being used. Are we sure that end users of online services have really consented to their data being used to construct, for example, weapons systems or systems which could be used to suppress free speech?

8. All AI systems developed so far, and in particular those developed using ML techniques, are inherently confined to a single domain of "intelligence." (This, by the way, is why concerns about "consciousness" or "sentience" in AI systems seem to me to be a distant worry rather than an imminent concern: at present, nobody in the field has a plausible route towards the development of a generalised intelligent system. The best chess playing system in the world is incapable of playing even a simpler game such as draughts. It seems unlikely that the most pressing issue facing society will be an AI system that develops an interest in real-world political domination.)
9. But even within these limited domains, ML approaches are incapable of escaping the bias of the training data they are trained with. Human failings, such as racism, have repeatedly been replicated in ML systems trained on real-world data. And because the rules produced by ML approaches are incomprehensible to humans, they are generally impossible to modify to remove their learned biases. For example, when a Microsoft Research chatbot became offensive, it was not modified and instead was simply shut down. For more subtly biased systems, this might be an unattractive option to companies (or indeed governments) and so ML-based systems might end up perpetuating the same systemic biases that modern societies are striving to remove from human interactions.
10. What is more, not only are the rules of a particular ML system impossible for a human to understand, but its decisions are impossible to explain. Consider an AI weapon that decides to kill a person based on their perceived threat, or an AI chatbot that issues a racial slur. In both cases, the rules that the system has learnt are in one sense objective (precisely the same outcome will take place if the same input could be presented to the system a second time), but it is impossible to discover the machine's reasoning in a human-comprehensible form. The ethical basis of the machine's decision is inaccessible: why, exactly, was the deceased person thought to be a terrorist? What basis did the car have for driving off the road? In some systems, there might be an overlay of high-level human-written rules that can be understood and explained, but the legal and ethical basis for the whole system's decisions will never be explicable. This creates both legal uncertainty (where does legally-enforceable blame lie?) as well as ethical concern (are we comfortable as a society to entrust these decisions to a black box which we will never be able to interrogate or modify?).

11. Finally, ML systems can be vulnerable to adversarial inputs. Humans can study the behaviour of a ML system, and in some instances can then create special-purpose input data which fools the ML system into behaving in undesirable ways. For example, an ill-intentioned person might display a printed picture to a self-driving car with the result that the car crashes. Or someone might craft internet traffic that gives an automated weapons system the impression of a threat, resulting in an innocent person's death. Of course, both of these are possible with non-ML systems too (or indeed with human decision-makers), but with non-ML approaches the reasoning involved can be interrogated, recovered, and debugged. This is not possible with many ML systems.
12. Many other forms of AI system exist, although these are not often perceived in popular imagination as AI (in fact, it is a commonplace in AI research that once a particular problem has an effective technical solution, it stops being "Artificial Intelligence" in the public eye). There are many AI systems in weapons already — for example, no modern anti-missile system would be possible without the use of AI systems. These other forms of AI system are much less effective at tasks such as recognising faces or driving cars, but have significant ethical advantages to society because they do not rely on mass invasion of privacy in their development, are not inherently subject to replication of human biases, can be understood and modified, and produce decisions which can be legally and ethically explained after the fact.
13. I suggest that as a society we have a long-term interest in encouraging technical approaches to AI which are amenable to ethical and legal oversight and explanation, even though there are significant short-term benefits from the approaches which are currently technically dominant. Even though Machine Learning has led the recent step change in AI system capabilities, and ML systems can offer great benefits to society, in my view it is also vital for public policy to foster long-term research into other kinds of AI, and to carefully regulate the development and uses of ML systems. Regulation could helpfully focus on issues of user privacy and consent around datasets used for training, and on identifying and limiting the legal and ethical effects of black-box systems.

6 September 2017

Richard Ebley – Written evidence (AIC0026)

Submission of evidence to the Select Committee on Artificial Intelligence

1. All levels of government and all public funded organisations need to demonstrate good management if AI is to succeed.
2. I suggest ISO 9001 is used to achieve this.

24 August 2017

The Economic Singularity Supper Club – Written evidence (AIC0058)

House of Lords Select Committee submission

Submission on behalf of the Economic Singularity Supper Club (ESC). The signatories are listed at the end.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

AI is our most powerful technology and it already produces impressive results. For instance it beats humans at many image recognition tests, and it is superhuman at games like chess and Go. But it is improving at an exponential rate, and in many ways we are only at the beginning of the AI story.

Moore's Law (the observation that computers double their performance per pound every 18 months or so) is morphing, not halting, and progress may even accelerate as new chip architectures and new approaches to computing (like quantum computing) are adopted.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

Yes. AI will bring great benefits, and also significant risks, which must be thought about and managed.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

(In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.)

The Economic Singularity Supper Club was established to raise awareness of the possibility of technological unemployment, and to promote activities which might help produce effective responses to it.

There are many aspects of AI that require consideration, but we believe that technological unemployment is the one that will have the greatest impact on humanity (at least until and unless we develop artificial general intelligence), and

we believe it is not currently receiving the detailed, concerted study that it requires. We argue the following:

1. In the coming decades, AI and related technologies will have enormous impacts on the job market. At the moment, no-one can predict exactly what will happen or when.
2. The outcomes could be anywhere from very good to very bad, and which one(s) we get may depend significantly on the actions taken (and not taken) by governments and others in the coming few years.
3. One possible impact is that many people will be unemployable within a relatively short space of time.
4. Responsible social leaders must address the question of how our economies could adapt to cope with the fast-approaching and unprecedented changes.
5. This includes addressing seriously issues like income replacement and resource distribution.
6. The solutions are complex and not obvious: serious work should start soon, including detailed scenario planning, modelling, large scale experiments, and communications plans.
7. This work could be sponsored by (and located in) government department(s), the tech giants, the existential risk organisations, other think tanks and universities.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

At different times and in different places there will undoubtedly be those who benefit relatively more and others who benefit relatively less. During the 2020s, for instance, self-driving vehicles will displace large numbers of human drivers. But the impact will be broader than many people expect. It is often argued that people doing repetitive jobs are the most vulnerable, but smart machines are already showing themselves capable of more than rote, repetitive tasks.

But overall, if we manage the transition to an AI-centric world, we will all benefit tremendously.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Yes. AI is going to influence all our lives intimately and comprehensively, so we should all have some understanding of what it is, and also what it is not. Plenty of material is already available, and more is being created all the time. People in positions of influence – including politicians, business leaders, journalists etc. –

have a particular duty to inform themselves, and not to initiate or perpetuate misunderstandings.

Because the solutions are complex and not obvious, the most urgent requirement is for serious analytical work to start soon, including detailed scenario planning, modelling, large scale experiments, and communications plans.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

(In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.)

Andrew Ng (formerly a senior researcher at both Google and Baidu) says: “Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don’t think AI will transform in the next several years.”

7. How can the data-based monopolies of some large corporations, and the ‘winner-takes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

There are network effects in information-based industries which can favour the emergence of monopolies. But there is also fierce competition, and business models change quickly, creating losers out of winners and vice versa. The histories of IBM, Microsoft, and Apple illustrate this clearly, and Google, Amazon, Facebook are not immune. Regulation may be required at times, but because it tends to tackle issues which have already faded, it can have modest or even negative impacts, and should therefore be embarked upon with great caution.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

(In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.)

See our answer to question 3.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable? When should it not be permissible?

See our answer to question 3.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

If and when technological unemployment arrives there are many different possible outcomes, ranging from the very good to the very bad. As we said before, serious analytical work on this should start very soon, including detailed scenario planning, modelling, large scale experiments, and communications plans. Government should encourage this and perhaps sponsor it, and pay close attention to the findings.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

Although the UK has a thriving and impressive AI research base and startup community, it cannot be denied that the leading AI development centres are in the US and China. This is not necessarily a problem: AI is a global phenomenon, and if we are to navigate successfully to an AI-centric world we will do it globally. Interested parties in the UK should be working closely with those elsewhere to make sure this happens.

Signatories (in alphabetic order):

Calum Chace, writer

Charles Radclyffe, head of technology at Deutsche Bank Labs

Daniel Hulme, Founder and CEO, Satalia

Julia Begbie, director, KLC School of Design

Mark Chadwick, CEO of Carbon Clear Limited

Radhika Chadwick, senior partner, Ernst & Young

Will Gilpin, programme manager, Schibsted Media

Other members of the Economic Singularity Supper Club (ESC) are unable to sign this submission because of their employers' policies, or because they are making individual submissions.

3 September 2017

Professor Lilian Edwards – Written evidence (AIC0161)

Lilian Edwards
Professor of E-Governance
University of Strathclyde
6 September 2017

I am only addressing ETHICS, points 8 and 9 and primarily looking at this as a lawyer not an ethicist.

1. Key problems described in the field relating to algorithmic governance

1.1 Discrimination and unfairness

A great deal of the extensive recent literature on algorithmic governance has wrestled with the problems of discrimination and fairness in ML⁴⁴². Most problems of bias and discrimination in ML systems arise from biases in the training set data used to build them. These correlations frequently relate to “protected characteristics”, a varying list of attributes about an individual such as so-called race, gender, pregnancy status, religion, sexuality and disability, which in many jurisdictions are not allowed to directly (and sometimes indirectly⁴⁴³) play a part in decision-making processes. Algorithmic systems trained on past biased data without careful consideration are inherently likely to recreate or even exacerbate discrimination seen in past decision-making. For example, a CV filtering system based only on past success rates for job applicants will likely encode and replicate some of the biases exhibited by those filtering CVs or awarding positions manually in the past. While some worry that these systems will formalise explicit bias of the developers, mostly these systems will be built with indirectly, unintentional and unknowing bias. Problems also arise from the subjective ways we collect training set data. We never have a complete picture in the data we collect. Most data are not gathered at random from society, but collected in ways that can be problematically skewed. For example, we cannot measure who breaks the law — only who is convicted of it.

One way forward is to try to build fair or non-discriminatory ML systems where these characteristics are not explicitly fed into the system, even if they have some predictive value — e.g. by omitting the data column containing race or gender. But this may have the downside that these systems perform less well than less “fair” systems; or that they recreate from various inputs the offending bias.

It is worth noting that in the EU, there have been far fewer scare revelations of “racially biased” algorithms than in the US. While some of this may be attributed

⁴⁴² See the useful survey in Brent Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, 3 *BIG DATA & SOCIETY* 2 (2017), especially at section 7.

⁴⁴³ In relation to the U.K., see Equality Act 2010, c. 15, s. 19; for discussion of US law, cf. Solon Barocas and Andrew Selbst “Big Data’s Disparate Impact” (2016) 104 *CALIFORNIA LAW REVIEW* 671.

to a less investigative journalistic, civil society or security research community, or conceivably, a slower route towards automation of state functions, it may also simply reflect a less starkly institutionally racist mass of training data.

Not all problematic correlations that arise in an ML system relate to characteristics protected by law. This takes us to the issue of *unfairness* rather than simply discrimination. As an example, is it fair to judge an individual's suitability for a job based on the web browser they use when applying, for example, even if it has been shown to be predictively useful?⁴⁴⁴ In the European DP regime, fairness is an overarching obligation when data is collected and processed⁴⁴⁵ something which is sometimes overshadowed by the focus on legitimacy and particular user rights

1.2 Opacity and transparency

Users have long been disturbed at the idea that machines might make decisions for them, which they could not understand or countermand. In the world before ML systems, such worries were at least partially met by subject access rights (SARs) in European DP law, with some equivalents in US law in domains such as credit scoring. In the public sphere, notions of transparency gave us widespread rights of freedom of information (FOI) or more recently "open data". In Europe a specific though rather under-used right also appeared in the Data Protection Directive 1995 art 15 to stop a decision being made solely on the basis of automated processing.⁴⁴⁶ Data subjects had a right to obtain human intervention (a "human in the loop"), in order to express their point of view but this right did not, notably, contain an express right to an explanation (see below) . This right was updated in the GDPR art 22 to extend to a more general concept of decision making that included profiling⁴⁴⁷.

ML algorithms, unlike their predecessors, rule based systems, are not built for transparency to humans but for performance. They are often described as "black boxes" or as non-interpretable. Designers of ML systems formalise a supervised or unsupervised learning approach as a learning algorithm. This software is then run over historical training data. At various stages, designers usually use parts of this training data that the process has not yet "seen" to test its ability to predict, and refine the process on the basis of its performance. At the end of this process,

⁴⁴⁴ *How might your choice of browser affect your job prospects?*, THE ECONOMIST (Apr. 11, 2013).

⁴⁴⁵ GDPR, art 5(1)(a).

⁴⁴⁶ This is interestingly interpreted by Jones to imply that European systems are more interested in the human dignity of data subjects than the US system: see Meg Leta (Ambrose) Jones, *Right to a Human in the Loop: Political Constructions of Computer Automation & Personhood from Data Banks to Algorithms*, 47 SOCIAL STUDIES OF SCIENCE 216 (2017).

⁴⁴⁷ GDPR, art 4 (4). "Profiling" includes "any form of automated processing of PD consisting of the use of PD to evaluate certain personal aspects relating to a natural person in particular to analyse or predict [...] Performance at work, economic situation, health, personal preferences, interests, reliability or behaviour, location or movements". Note such profiling may be achieved other than by ML; see discussion in section 0.

a *model* has been created, which can be queried with input data, usually for predictive purposes. Because the logic of these ML models was induced, they can be complex and incomprehensible to humans. This represents a fundamental challenge to transparency, and has exacerbated the increasingly panicky claims of bias, unfairness and discrimination discussed above.

2 The “right to an explanation”?

In response there has been a flurry of interest in law in a so-called ‘right to an explanation’ that has been claimed to have been introduced in the EU General Data Protection Regulation (GDPR). However as already noted a similar remedy had existed in the EU Data Protection Directive (DPD), since 1995. Technologists especially have seized on the right to an explanation as a way forward to gain consumer trust. Yet curious serious problems with the alleged right to an explanation are canvassed in L Edwards and M Veale “Slave to the Algorithm?” forthcoming Duke University Law and Technology Review, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855 . These include:

- Art 22 applies only to systems where decisions are made in a “solely automated” way – i.e. no human in the loop - and there are very few of these and fewer that are “significant” (see below).
- Art 22 applies only to a decision that produces legal or other significant effects. This is vague in the extreme. Is a system advising buying choices or targeting adverts significant?
- Art 15 which provides a right to “meaningful information about the logic involved” in a decision making system may be more useful. But there is an unresolved doubt about whether this only applies to information available before the system make a decision about a particular data subject (see Wachter et al) which so far cripples this right with uncertainty/
- Perhaps most importantly arts 15 and 22 of the GDPR both only apply where decisions are made based on personal data. Yet algorithmic decisions which affect people may not involve personal data. The most obvious example is self driving cars. They may kill people – drivers or pedestrians – yet the data processed may be entirely related to traffic, road conditions and other non personal matters.
- Other circumstances may involve data that was once personal but has been allegedly anonymised. Again this raises huge amounts of doubt as there is little way to tell when data is capable of being reidentified or “repersonalised”

In short the academic debate and popular furore about a specifically **GDPR** right to explanation is distracting. **What would be better would be to consider what a useful new consumer protection rule giving basic rights to transparency when algorithmic decisions are made would look like.** This debate is happening elsewhere (often *because* there is no local DP law as in the US) e.g. in NY a Bill has just been drafted for municipal transparency, see <https://www.nytimes.com/2017/08/24/nyregion/showing-the-algorithms-behind-new-york-city-services.html?mcubz=3> . We should learn from these efforts

elsewhere while at the same time incorporating some of the ideas of fairness and informational transparency that come through DP law.

3 Transparency fallacies: transparency is not the whole, or final, answer

There is an enormous debate in ML about whether technology can in fact produce “meaningful explanations”. I leave this to my CS colleagues; though there is certainly a desperate need to test how different types of explanations, generated from different datasets, used to make different decisions for different types of data subjects may work to improve comprehension and trust of users. This needs a programme of funded research; but more than anything it also needs willingness from public bodies making real decisions to find ways to allow researchers to test using these datasets. There are clear problems of privacy and confidentiality here which will need well discussed in advance.

Transparency even if practical merely gives a user a better idea of how a decision was made. It may not even tell the user how that decision was made specifically for him or her (“global” rather than “local” explanations). It will certainly not necessarily give them the tools *to combat* that decision, and to imagine it will, alone, may do more harm than good.

Existing imbalances in power; consumer fear and inertia in the legal market; lack of legal aid and advice; lack of funding and title to use for consumer organisations and citizen rights groups ; and bureaucratic ways to resist accountability (as seen with FOI); will persist. In short, transparency may be a band aid and at worst a replacement for actual solutions for users harmed by biased or erroneous algorithmic systems. The history of privacy shows that increased transparency – in the form of longer and longer privacy policies – with alleged control over data collection by consent – does not produce more privacy or more trust by users.

The real answer to building a better algorithmic society is probably not a “right to an explanation” – though it won’t hurt - but to build better systems using better data. This is as much a matter of money and political/ commercial will as it is technical or ethical. Very few companies or governments currently set out to build biased or unfair algorithms but it is often the easiest and cheapest thing to do using existing training data infused with existing bias and partial sampling, and without doing preparatory work to avoid this.

4 Things we could do with the GDPR to build better algorithmic systems

Several novel provisions in the GDPR try to provide a societal framework for better privacy practices and design:

- **requirements for Data Protection Impact Assessments (DPIAs) and privacy by design (PbD).** DPIAs are supported in many ways by the Information Commissioners Office but have yet to trickle down as normal to the private sector. In particular given guidance already out it is likely every new ML system may be regarded as risky enough to need a PIA. We need urgently to study how these can be folded into commercial development time cycles and profit motivations, and not just become tick-box bureaucracy.
- **non-mandatory privacy seals**
- **non mandatory certification schemes** showing compliance with aspects of the GDPR.

These provisions may help produce both more useful *and* more explicable ML systems. Of course these GDPR based remedies will again not be applicable if the system is built using non personal data.

5 Things beyond the GDPR we could try to fix using law

One of the key issues reducing both user trust in algorithmic systems and the ability of external parties (like journalists, scientist and lawyers) to audit such systems for unfairness and bias is that **many such systems are proprietary secrets and protected by legal protection of some kind – possibly IP but most often trade secrets.**

Assuming the UK implements the new EU Trade Secrets Directive we may have an opportunity when implementing it to carve out a public interest exception allowing “reverse engineering” or even actual disclosure of trade secret algorithms. However notably such disclosures would have to include training set data as well as the actual algorithms or models (which are often fairly well known already)

In **public ML systems such as bail or probation systems** (which have created the most notorious results in the US) there must be a case for such systems to be built as far as possible as open scrutable datasets and algorithms.

6 September 2017

Electronic Frontier Foundation – Written evidence (AIC0199)

Comments of Electronic Frontier Foundation September 6, 2017

Peter Eckersley, Ph.D.; Jeremy Gillula, Ph.D.; Jamie Williams
Electronic Frontier Foundation

1) The Electronic Frontier Foundation (EFF) submits the following comments in response to the House of Lords Select Committee on Artificial Intelligence's Call for Evidence, available at <http://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/Artificial-Intelligence-call-for-evidence.pdf>. EFF is a member-supported, nonprofit, public interest organization composed of activists, lawyers, and technologists, all dedicated to protecting privacy, civil liberties, and innovation in the digital age. Founded in 1990, EFF represents tens of thousands of dues-paying members, including consumers, hobbyists, computer programmers, entrepreneurs, students, teachers, and researchers. EFF and its members are united in their commitment to ensuring that new technologies are not used to undermine privacy and security.

2) For the purposes of our comments we use a fairly broad definition of AI, which includes everything from simple machine-learning (ML) systems to advanced deep-learning techniques. While others may focus their remarks on more advanced systems, we believe it is important to acknowledge that even simple AI systems in use today (which some may no longer even classify as AI) are already having a dramatic impact on society.

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

2. Is the current level of excitement which surrounds artificial intelligence warranted?

3) We have made some initial studies of the pace of technical progress with our AI Progress Measurement initiative (available at <https://www.eff.org/ai/metrics>), which surveys problems, metrics, and benchmarks from the machine learning research literature, and tracks progress on them.

4) Any prediction of future technological development is prone to significant limitations and methodological difficulties, so we wish to stress that the pace of technical progress over the next 5, 10, and 20 years is of course highly

uncertain. However, the data we have collected shows that this field is making rapid advances on a very wide range of problems. Influxes of talent, resources, and computing power will likely continue this trend.

5) Although there remain many daunting obstacles and difficult tasks that AI is not yet close to solving, there is evidence to support claims that machine learning could have significant economic impacts in a growing number of domains over the next 20 years, and that there is some possibility of drastically transformative AI technologies emerging in the next 10-30 years.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

6) This question combines several different issues. Winner-takes-all monopolies are not a new phenomenon in the computing industry, which often exhibits strong economies of scale, lock-in effects and even stronger network externalities, where the usefulness of a product is proportional to the number of people already using it. Many of the present concerns about big data and market power are extensions of these pre-existing and unsolved policy problems in the technology industry. We will not attempt to address them in this submission.

7) But there are also several competition policy questions that are quite specific to machine learning and AI research. These essentially derive from two pre-conditions: (1) present machine learning techniques require an enormous number of examples to successfully learn things; and (2) typically, only large technology companies are in possession of enough examples--i.e., the photos, emails, text messages, location histories, and/or sensor feeds of hundreds of millions of people--necessary to conduct machine learning research. This has given large companies a significant advantage when conducting basic research on certain machine learning problems.

8) Fortunately, this lead is not universal across all machine learning problems. For certain tasks, the academic community has been able to build its own large datasets that are comparable to privately held ones, or at least sufficiently large enough to achieve research breakthroughs. In other cases, technology companies have voluntarily shared data in order to stimulate open research on problems they consider important, and/or to promote themselves to prospective employees.

9) But for other tasks, the process of sharing datasets is somewhat complicated by the private and sensitive nature of the data required. For example, if one wishes to use machine learning to understand and process email better (for instance, in making a better spam filter or in making an agent that can read and handle some types of email for you), one needs large datasets

made from genuinely representative sample emails. Since it is inherently problematic to share large, representative datasets of people’s private email, only major email providers can directly perform this sort of machine learning research. Similar problems apply to research based on network traffic data, server logs used for cybersecurity purposes, patterns of online behavioural data, and many other categories.

10) In the long run, we are unsure how serious a problem this will be. New algorithmic techniques such as federated learning⁴⁴⁸ and differential privacy⁴⁴⁹ could in theory allow more sharing of privacy-sensitive training data, but it is doubtful they could fully close the productivity gap between AI researchers working for the largest tech companies and the rest of the research community.

11) Though these effects are likely to continue to confer advantages to established players in the race to apply machine learning to some economically important tasks, it is less clear that there are sound competition policy interventions available in the AI space specifically at this time. Probably the most constructive role that governments could presently play is to provide additional incentives and support for the creation of open research datasets, particularly where there are algorithmic ways to solve the privacy and security problems that would otherwise hamper the use of that data for research purposes.

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

12) The ethics, safety, privacy, and social policy questions raised by existing machine learning technologies are quite complicated, and they will grow much more so as artificial intelligence becomes more autonomous and capable of more diverse and learned forms of action. There is an emerging field of academic and industry research focused on these questions, but they are far from solved.

13) In one pressing short-term example, the use of machine learning and other statistical algorithms is already disparately and unfairly impacting different populations in society due to problems that include not only biased training datasets but also the emulation of human prejudices as a result of a statistical problem called *omitted variable bias*. Omitted variable bias occurs when an algorithm lacks sufficient input information to make a truly informed prediction about someone, and learns instead to rely on available but inadequate proxy variables. For instance, if a system was asked to predict a person’s future educational achievement, but lacked input information that captured their

⁴⁴⁸ See B. McMahan et al, Communication-Efficient Learning of Deep Networks from Decentralized Data, AISTATS 2017, <https://arxiv.org/abs/1602.05629>.

⁴⁴⁹ See C. Dwork 2014, Algorithmic Foundations of Differential Privacy <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>

intelligence, studiousness, persistence, or access to supportive resources, it might learn to use their postal code as a proxy variable for these things. The results would be manifestly unfair to intelligent, studious, persistent people who happened to live in poorer areas.

14) Detecting and analyzing this sort of unfair impact is complicated by the fact that, depending on the context, it is not always clear what the appropriate measure of bias should be--or in other words, what results would be "fair" and what results would be "unfair." The topic of "algorithmic fairness" has spawned an entire research field, some aspects of which we discuss in our answer to question 10 below.

15) The current deployment of machine learning techniques poses a serious privacy risk. Specifically, machine learning has enabled efficient large-scale surveillance both by intelligence agencies and commercial actors. In the past, large-scale surveillance of a population was limited by the human resources available to sift through the data collected. Only societies like East Germany, that were willing to recruit one informant per 6.5 citizens, could possibly watch and pay attention to all of their citizens' actions. But the combination of already-deployed surveillance technologies and machine learning for analysing the data will mean that exhaustive surveillance is becoming possible without the need for such enormous commitments of money and labour. The potential of machine learning to enable such effective large-scale surveillance has reduced the price tag of authoritarianism, and poses a novel threat to free and open societies. For this reason, EFF believes that machine learning algorithms should only get access to a person's data with their consent and control, or a properly issued warrant.

16) More broadly, when algorithms make decisions that affect human lives in ways that may be mundane but expensive (e.g., price discrimination) or profound (e.g., sentencing or bail recommendations, use of machine learning in mass surveillance), there must be transparency, openness, due legal process, and accountability for intended and unintended consequences, as we describe in our response to the next Question 9.

17) In the medium term, the impact of AI on labour markets deserves serious attention. The EFF has no view on the right solution to that problem, but we do think there is some risk of the market providing many fewer well-remunerated jobs in the coming decades, or many fewer jobs in general, and that planning in advance for this possibility is an important task for societies that are presently fundamentally motivated by and organised around the Protestant work ethic. We would urge those across the ideological spectrum to think seriously about what kinds of society they would want to see if less human labour was practically necessary for prosperity. Would we be comfortable with fewer people working, and willing to share resources with them? Would it be better to create new forms of artificial work? How will we preserve the sense of opportunity, self-worth and status of humans in such societies?

18) In the longer term, the ethical and societal questions around AI may be even stranger and more profound. What would it mean for humanity to share the planet with other types of intelligence? Though entertaining, such topics are extremely speculative and at present more usefully addressed by academic research and science fiction than by concrete policy making, though we do think that there are some exceptions. For instance, if highly transformative artificial intelligence technologies were developed in the future, the risks associated with computer insecurity would almost certainly rise dramatically. As a result, we believe the possibility of AI advances in coming decades are a reason to increase funding and incentives for the creation of secure computing infrastructure and effective defensive cybersecurity systems today.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

19) Any AI systems that significantly impact the rights, freedoms, or lives of large populations of people must at least be auditable, if not transparent.

Examples of systems where some level of transparency is necessary include:

- AI systems used for government purposes (e.g., to advise judicial decisions, to help decide what public benefits people do or do not receive, and especially any AI systems used for law enforcement purposes);
- AI systems used by companies to decide which individuals to do business with and how much to charge them (e.g., systems that assign credit scores or other financial risk scores or financial profiles to people, systems that advise insurance companies about the risk associated with a potential customer, and systems that adjust pricing on a per-customer basis based on the traits or behavior of that customer);
- AI systems used by companies to analyze potential employees; and
- AI systems used by large corporations to decide what information to display to users (e.g., search engines, AI systems used to decide what news articles or other items of interest to show someone online--if they make those decisions based on individual user characteristics--and AI systems used to decide what online ads to show someone).

20) The appropriate level of transparency will be different for each of the scenarios described above. For example, given the tremendous impact AI systems used for government purposes or financial decisions can have on people's lives, such AI systems should be completely transparent--regardless of whether or not they were publicly or privately developed. The public and all those potentially impacted should have access to the algorithms (and as we describe below, training data), and the systems should be subject to regular, published audits, which should include measuring how the system performs under various fairness metrics, to ensure that the system continues to function as expected (and is not causing any discriminatory, unfair, or unintended

effects). These audits could be performed by the organization responsible for the AI system or an independent governmental body, but they must be mandated to ensure that they are performed in a regular and timely manner. And of course, all audit results should be immediately made public.

21) A lower level of transparency may be appropriate for algorithms that have a lesser impact on people's lives, such as search engines and news feed algorithms (although the impact of these algorithms is by no means insubstantial). These algorithms are often closely guarded trade secrets that required tremendous R&D expenditure to get right. As such, a lower level of transparency, such as effective auditability for discriminatory outcomes (i.e., the ability of independent parties to test the system to ensure it doesn't unintentionally discriminate based on characteristics like race, religion, etc.) without complete transparency (i.e. publication of the algorithm) might be sufficient to protect the public interest, particularly if individuals or organizations who wish to access the APIs to audit the systems must first sign agreements not to use any data they derive during the course of the audit for competitive purposes.

22) Additionally, when it comes to AI systems, transparency should not be limited to just the algorithm or the code running the system; the datasets used to train an AI system are just as critical in order to ensure transparency. This is because datasets can have a tremendous impact on the performance of the AI system, causing problems even if the algorithm itself is flawless and unbiased.⁴⁵⁰ Further, knowing what datasets were used to train an AI can help independent auditors discover where an AI system might be functioning in a biased or unfair manner.

23) As an aside, we close by noting that the dangers of black-box systems apply just as much to non-AI systems as to AI systems. A sentencing algorithm or a credit score doesn't have to use convolutional neural networks or other deep learning techniques in order to have a discriminatory or otherwise unfair impact on people's lives.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

⁴⁵⁰ For example, see *Google Photos labeled black people 'gorillas'*, USA Today, <https://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/> for an example where a dataset that underrepresented black people resulted in unintentional results, or *Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day*, The Verge, <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>, for an example where an otherwise reasonable social AI system learned to be antisocial based on its interaction with Twitter users.

24) In general, the Electronic Frontier Foundation is skeptical about government regulation of technology. Legislation often has unintended consequences and can easily interfere with the process of innovation. Most aspects of artificial intelligence are far too speculative and immature to be appropriate subjects for regulatory action at present. However there are some well-documented and serious problems with currently deployed machine learning systems for prediction and classification in institutional decision making.⁴⁵¹ They are urgent enough that careful, judicious, and consciously experimental regulation may be warranted in some domains.

25) For example, the problems of racial, gender, and other forms of demographic bias in deployed machine learning systems are so severe that they constitute serious public policy problems. Regulation requiring processes to at least measure and report, if not specifically correct,⁴⁵² such biases should be considered in cases where the decisions, choices, or recommendations these systems make significantly impact people's lives.⁴⁵³

26) On the topic of transparency and explainability, the EU is conducting a significant regulatory experiment with the "right to explanation" contained in the GDPR. Providing good explanations of what machine learning systems are doing is an open research question; in cases where those systems are complex neural networks, we don't yet know what the trade-offs between accurate prediction and accurate explanation of predictions will look like.

27) In some domains of application, there are fundamental reasons for optimism about those trade-offs and therefore about the GDPR's rules. For instance, where a classification system is trained on a set of readily describable

⁴⁵¹ To our knowledge, the real and well-demonstrated practical problems currently exist in comparatively simple systems such as regression models, rather than complex neural networks that would more fairly deserve the label "artificial intelligence". However it would be prudent to craft any regulatory principles so that they could apply to either type of system.

⁴⁵² Several methods for correcting bias have been proposed in the literature. For an entry to this literature, see M. Hardt, E. Price and N. Srebro, Equality of Opportunity in Supervised Learning, *NIPS 2016*, <http://papers.nips.cc/paper/6373-equality-of-opportunity-in-supervised-learning>, and <https://research.google.com/bigpicture/attacking-discrimination-in-ml/>. Since there are several incompatible standards for what fairness and de-biasing might mean (see J. Kleinberg, S. Mullainathan and M. Raghavan, Inherent Trade-Offs in the Fair Determination of Risk Scores, *ITCS 2017* <https://arxiv.org/abs/1609.05807>) it may be prudent to require organisations deploying machine learning for high-stakes decision making to select and justify one standard while measuring and reporting their rates of deviation from the others. We would caution however that "maximising the accuracy of predictions" is rarely an appropriate notion of fairness. We would caution also that these bias mitigation techniques do not address the problems of building models from inherently biased data sources, and good regulation would find ways to incentivise companies to be skeptical of their training data and find ways to improve on or work around its flaws.

⁴⁵³ We have no specific recommendations about how to delineate between high and low-stakes applications of machine learning, but rules that apply when decisions have an expected monetary value above some threshold are one obvious way to achieve this.

input variables, it should be possible to provide good statistical explanations even if the classifier is a very complex neural network.⁴⁵⁴ In other domains, particularly when the inputs are complex data like images or video, it's not clear if accurate simple explanations of predictions will always be available, or what trade-offs will have to be made to obtain them.

28) Given these technical uncertainties, the EU's "right to explanation" should be viewed as a regulatory experiment, and its successes and failures should be continually evaluated. To the extent that Brexit gives the UK additional flexibility in adopting the GDPR or otherwise, we would ask the question, "how can the UK take a position which builds on the utility of the EU's experiment?" That might mean adopting clearer and stronger incentives for explainability, if the GDPR rules appear to be bearing fruit in terms of high-quality explanatory technologies, or it might mean moving in the direction of different types of rules for explainability, if that technical research program appears unsuccessful. Incorporating the EU's own reviews and reforms would also be important.

29) Finally, we caution the government from enacting any regulations that are narrowly tailored to specific AI techniques. The right course when it comes to regulating AI is to focus on each of the AI system's impact and domain of application, as opposed to the underlying technical methods. Regulations that focus on the technology, instead of the impact, will likely fail to protect the public, and they may also threaten innovation. As we noted in our answer to Question 9, a sentencing algorithm or a credit score doesn't have to use convolutional neural networks or other deep learning techniques in order to have a discriminatory or otherwise unfair impact on people's lives. Innovation in the field of AI is proceeding rapidly--both in terms of the capabilities of AI systems and the scope of problems they can solve. But the field of AI safety research is also growing--both in terms of ensuring that AI systems act as expected and also, as we mentioned earlier, in terms of ensuring that AI systems act fairly. We urge the Government to focus on the end result of the deployment of AI systems when determining what regulations are appropriate.

6 September 2017

⁴⁵⁴ For instance, the statistical explanation for a low insurance premium could be "In cases like yours, our model found complex relationships between age, gender and driving habits that predicted risk of an accident. Generally speaking, it helped that you were younger, female, that you started driving at a younger age, and that you rarely drove in high-risk locations."

Dr Julian Estevez – Written evidence (AIC0021)

The term *artificial intelligence* (AI) was coined in 1956 by John McCarthy. However, it is in the last 15 years when this science became so popular. In my opinion, three key factors accelerated this phenomenon:

- Some authors point out that progress in computer technology was one of the most important factors, but I personally find that the understanding of the algorithms that rule AI was of higher importance. As [1]⁴⁵⁵ points out, hardware development is easier to measure rather than the understanding of the mathematics and algorithms of AI.
- Big data repositories that exist since only the last decade. This huge amount of data lead to the development of data-driven AI, commonly known as machine-learning. They permit the mass analysis of images, sensors data during long periods, personal data classification...
- The tools that giant companies made public, such as Google's Tensorflow.

However, in my opinion, the technology is still in its infancy in terms of its applications in society and industry. More research is necessary so that a product becomes full available to the public and guarantees robustness, safety and usage of intelligence.

As far as I'm concerned, I define AI as a series of rules consisting on mathematical operations with different variables that create what we call "learning", usually very specific tasks (for instance, pattern recognition, items classification, face recognition, image analysis, data search, etc).

Despite this apparent simplicity of AI, it is enough for both adverse consequences of this technology: on the one hand, the threatening of millions of jobs and fear of this technology to overwhelm us. On the other hand, the development of some applications that make our lives easier and will transform economy and society. I am talking about of the advance in search engines, the improvement in medical image diagnosis, driverless cars or the progressive equipment of capabilities for robots.

In this document, I would like to point out some worries of myself related to the future massive applications and implementation of AI systems in factories and society. Some impacts of this science are being properly forecasted, such as the probable unemployment and the enhancement of society services, but I consider that next points have not been argued enough yet.

⁴⁵⁵ Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global catastrophic risks*, 1(303), 184.

- 1- Need of blackbox AI regulation and prevention of data usage for discrimination
- 2- Bentham's Panopticon in industries
- 3- Adversarial machine learning

1- Need of blackbox AI regulation and prevention of data usage for discrimination

Regulation of AI is one of the main concerns among researchers, public leaders and mass media.

The White House Office of Science and Technology Policy (OSTP) report⁴⁵⁶ discusses the issues of fairness and transparency, and brings up two different concerns:

- The need to prevent automated systems from making decisions that discriminate against certain groups or individuals.
- The need for transparency in AI systems, in the form of an explanation for any decision.

However, it is not an easy point. Today, we are surrounded by technology which most people don't have idea at all about its working, such as mobile phones, planes, the Internet, or even cars. But the key factor of all these things is that they do what we ask them to do, and that safety substitutes the hunger of the scientific fundamentals understanding of these apparatus. On the contrary, AI will be a very recurrent tool for professionals to take decisions that will affect our lives, and in most occasions, it will not be straightforward for us the systems response.

The process today for a bank to decide whether we fulfil their conditions to get a loan from them is not transparent at all for us. Simply, the banker might end up the conversation stating that *the bosses had decided that*. Or even an employer doesn't need to explain us why he selected another candidate for the job. But it is important that if these decisions are taken with an intelligent system, the rules it follows should be explained or tested somehow. Otherwise, these AI solutions become *blackboxes*, and nobody can really access to know on what evidence the machine decides on us. The lack of transparency might lead to the *carte blanche* of the uncontrolled discrimination.

This claim of clarity is even more necessary in healthcare applications which upon some symptoms or some magnetic resonance images, intelligent systems detect a tumour or suggest a medical treatment. Think about the next situation as a hypothetical example: let's imagine an intelligent exoskeleton that learns

⁴⁵⁶ United States (2016) Executive Office of the President. Preparing for the future of artificial intelligence. Technical report, National Science and Technology Council, Washington D.C. 20502, October 2016.

our movements and adapts to us. If we step wrong downstairs and start falling, would the intelligent system develop an extra strength, even considering that this effort might be dangerous for its servomotors? How can we certificate a system considering this exception?

The level of detail might not reach to the total description of the algorithms, but some kind of tests and transparency are needed, as it exists for the cars or aeroplanes today. Codes that rule the electronics of cars or aeroplanes are not public, but these vehicles must pass specific intense tests to check that they will work properly under a wide spectrum of conditions.

Finally, it is crucial to guarantee the access to personal data of any individual to be heavily restrained. It will reach the day in which personalized diagnoses based on some own variables will provide much information about our instant and future health. This info should be just available to us, and not to the employer that might take the decision to fire someone considering future predicted diseases. And same for the case of why a bank reject our loan petition. That shouldn't condemn us not to receive money from any other financial institution.

This last part tells nothing different from what we consider as a common sense issue. But I think that in a future we will have to pay a special attention to all our personal data.

2- Jeremy Bentham's Panopticon at industries

I think that Bentham's idea is helpful for explaining the next issue, even it is based on late 18th century. This philosopher developed the concept of a permanent surveillance feeling and its psychological effects. We can present a parallelism between this concept and the automation that permits a permanent control of workers performance. It's already some years since a person is physically unnecessary for surveillance tasks in many plants, due to the multiple sensors and buttons that need to be checked by the workers. But this technology is based on 3rd Industrial Revolution. Now, intelligent systems permit a further surveillance approach.

An essential part in machine learning and AI is the pattern recognition for the data automatic classification and prediction. And so, there are many examples of AI solutions that permit measuring the effect of each worker in the production and analyse the number of defects, production rate, material loss, over many other variables.

I find that this kind of information should be guaranteed to be used for the enhancement of the production process and potential training for the workers, rather than as a firing excuse. Besides, one of the Panopticon's probable effects is the alienation and low motivation of the workers. These aspects lead me to think that occupational safety and health normative might need to be updated to

forthcoming technologies, which include AI, robots, production flexibility and some other aspects that are not analysed in this document.

3- Adversarial machine learning

Adversarial machine learning is a broad category of AI cybersecurity, consisting in the planned alteration of an image that cheats the detection system. These slight changes might not be visible for the human eye, but the computer interprets another result (such as identifying another person when the individual wears coloured glasses). Adversarial attacks can take different forms, including audio and perhaps even text. The existence of these phenomena was discovered independently by a number of teams in the early 2010s.

This technique can be used to bypass a system, or have substantial security implications for access systems, factories or robots. Articles [3]⁴⁵⁷ and [4]⁴⁵⁸ collect examples of adversarial machine learning examples which human-eye cannot detect.

Now, there are already plenty of ways to hack self-driving cars, for example, that don't rely on calculating complex perturbations. Nevertheless, adversarial learning has still its limitations and the attacker must know the code of the machine to be cheated, among other features.

Conclusions

The three artificial intelligence impacts on society and industry presented on this document try to prove that still research and tests on this technology are a necessary challenging objective. As I mentioned before, AI is still on its infancy, but it will reach the moment in which applications will become massive among consumers and workers. A regulation is needed for that moment that will guarantee safety and transparent usage of new systems.

We are approaching a point at which regulation could create more of a bottleneck than the development of technology itself, like self-driving cars and drones. It is the moment for public policy administrators, scientists, lawyers, doctors and all other stakeholders to join and develop the regulation and state objectives that will assure future generations' welfare and social rights.

22 August 2017

⁴⁵⁷ Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

⁴⁵⁸ Moosavi-Dezfooli, S. M., Fawzi, A., Fawzi, O., & Frossard, P. (2016). Universal adversarial perturbations. arXiv preprint arXiv:1610.08401.

euRobotics Topics Group on 'Ethical, Legal and Socio-economic issues' – Written evidence (AIC0189)

<http://www.pt-ai.org/TG-ELS/>

Vincent C. Müller (chair)

University of Leeds, Interdisciplinary Ethics Applied (IDEA) Centre

Q1: What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- 1) The current state of AI is much advanced over what it was in, say, the year 2000, and we expect this development to continue for a few decades to come. Having said that, AI uses mostly the same techniques as it did in 2000 and has not seen fundamental new insights. What drives AI progress is the massive increase in computing speed, data storage and available data – where speed increase essentially causes the increase in available storage and data. The computing speed that can be bought for a given cost has doubled ca. every 1 ½ years since 1970 (roughly “Moore’s law”). This “exponential growth” means that the computation we can buy for 100£ now is over 1000 times faster than what we got for the same amount in the year 2000 (and it will be 2000 times faster by early 2019, etc.). Other developments in computing have seen even stronger exponential growth, e.g. Internet or smartphone use. While it is clear that exponential growth must come to an end, it is generally expected to continue in the next decade or two.

Q2: Is the current level of excitement which surrounds artificial intelligence warranted?

- 2) Assuming that the current development continues in the next decade or two, we will see solutions to a kind of problems that lend themselves to existing “technical AI” – exploiting more speed and data. This means essentially problems that can be described with sufficient technical precision or problems that can be solved with large amounts of data, especially through techniques of “machine learning” that drive most of the current progress. Which problems are of this kind and which are not is very hard to say – it appears, for example, that autonomous driving can be solved with this kind of technique, though perhaps not for all situations that a human driver could handle. We know, however, that a very large set of problems can be handled in this way, including many we have not even thought of: This is why we expect continuous progress.
- 3) Problems that cannot be solved in this way include the development of general intelligence like that of humans, mammals or birds. Solving these would require going beyond “technical AI” into an AI that uses deep insights from the study of natural intelligence (cognitive science). For the same

reasons, the next two decades will not see robots that replace jobs on a massive scale.

- 4) We will also not see independent *agents* based on AI, i.e. systems that individually try to achieve goals and pursue these, even if they might be in contradiction to human goals. In the absence of general intelligence and agency, we will not see a 'rise of the robots' against their human masters – this will very likely remain forever in the realm of science fiction.

Q3: How can the general public best be prepared for more widespread use of artificial intelligence?

- 5) Concerning the *labour market*: Warnings like the notorious one that up to 47% of jobs can be automated within the next 20 years disclose neither an understanding of AI (see 1-4 above), nor an insight into the functioning of labour markets. Technological progress has been around since the First Industrial Revolution in the 19th Century, and the consensus among labour economists is that this has not caused a long-run increase in unemployment. This is because technology allows us to produce new and cheaper goods and services, creating economic growth and more jobs (both in existing sectors and in new ones) in the process. The Digital Revolution is expected to have much the same long-run economic impact, creating more and better jobs because it complements rather than replaces workers in doing mentally as well as physically demanding tasks. Also, new technologies can perform many tasks for which there is a lack of workers willing to do these tasks of for which we think human labor should be used scarcely – the 3 D, "dull, dirty, or dangerous".
- 6) However, there will be *adjustment costs* because of changes in the composition of employment. Individual workers do lose their jobs and see their skills become obsolete, even if new opportunities arise elsewhere. New technologies lead to changes in the organization of work, demanding more flexible work practices, regular on-the-job training, and more decentralized decision-making. These adjustments are governed by a range of factors, such as technology uptake (usually between 15 to 30 years to go from 10% to 90% adoption), as well as legal and institutional framework for training and regulation – this is where governmental regulation can have significant positive impact.

Q8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

- 7) Rapid developments in AI have the potential of undermining *human values*, esp. moral responsibility ("the robot did it!"), compassion, and human dignity. Robots may also undermine certain distinctions that we are fond of, e.g. human and non-human, and they may deceive humans. In addition, many ethical problems concern possible *negative consequences* on the well-being of humans (and other sentient beings). This includes safety at the workplace, de-humanisation of certain environments (such as health-care), and easier killing of humans in war. – Here the question is: Are the benefits of AI worth the risks?

8) So, the beginning of 'policy' for AI and robotics is that agents have *to uphold human values and maximise good consequences*. One main value issues seem to concern honesty, esp. non-deception of users and customers. Also, we need to retain the principle that humans are responsible for their actions, while robots and other machines are not. The main policy concerns are of consequences and risk, so the right action is to evaluate such consequences and analyse which course of action likely produces the most benefit. In this analysis, it is currently disputed, for example, whether robots in care or military robots maximise good consequences.

Q9: In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

9) Lack of transparency is inherent in several AI techniques, esp. in "big data" and in machine learning. It can generate problems of predictability, which means that systems must be tested and certified to a high degree if they are to be used in critical environments (where human lives are at risk). Such certification is needed to avoid "free riders" that may flood a market with products that are cheaper but transfer risk onto others (e.g. autonomous cars on pedestrians).

10) Furthermore, the "General Data Protection Regulation" (EU 679/2016) which will become UK law in 2018, foresees a "right to explanation" in several contract-relevant cases. In such cases (e.g. denial of a credit card by a bank), lack of transparency would thus not be permissible. The "informed consent" demanded in several privacy regulations is also problematic if the system that one consents to is not transparent.

Q10: What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

11) The area of civic liability is a concern for regulation: Can current liability law be maintained for AI and especially for cases of complex interaction of humans with artificial systems? The recent Resolution of the European Parliament [16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2013)] provides a useful framework for a legal debate. Human-machine cooperation will cause product liability rules and traditional tort law principles to overlap, which will cause high levels of uncertainty and litigation, delaying innovation. The inadequacies of existing rules may suggest to *radically replace a fault based rule with a risk-management approach* (based on absolute liability rules), meaning that the party who is better placed to minimize the cost is held liable and forced to take out insurance.

Definition of AI Used

12) Artificial intelligence is the technical discipline that develops intelligent computing machines. Intelligence is the result of abilities (e.g. moving, perception, learning, interaction, reasoning) that allow a given system to successfully pursue its goals to some degree.

About the author & the response

Dr. Vincent C. Müller, University Academic Fellow (University of Leeds) & Professor of Philosophy (Anatolia College/ACT)

Chair of the euRobotics topics group on 'ethical, legal and socio-economic issues' (euRobotics is the European Robotics Association – both for industry and for academics)

President of the European Society for Cognitive Systems (1000 members)

Organiser "Philosophy and Theory of Artificial Intelligence" conference series – the next event is 04-05. November, 2017 at the University of Leeds:

<http://www.pt-ai.org/2017/>

Main pertinent publications:

- Müller, Vincent C. (ed.), (forthcoming), *Oxford handbook of the philosophy of artificial intelligence* (New York: Oxford University Press).
- Müller, Vincent C. (ed.), (2016), *Risks of artificial intelligence* (London: Chapman & Hall - CRC Press).
- Müller, Vincent C. and Bostrom, Nick (2016), 'Future progress in artificial intelligence: A survey of expert opinion', in Vincent C. Müller (ed.), *Fundamental issues of artificial intelligence* (Synthese Library; Berlin: Springer), 553-70.

Some parts of this are based on a "Position Paper: Ethical, Legal and Socio-economic Issues in Robotics" that the topics group formulated in March 2017. That paper had been written by V. Müller in cooperation with *Dr. Andrea Bertolini* (legal), Scuola Superiore Sant'Anna and *Dr. Emilie Rademakers* (socio-economic), KU Leuven. See http://www.sophia.de/pdf/pdf_others/2017_TG-ELS_position_paper.pdf

The main literature source for robot and AI ethics is <https://philpapers.org/browse/robot-ethics> (edited by V. Müller) See there for introductory texts. See also the policy documents on <http://www.pt-ai.org/TG-ELS/policy>

6 September 2017

Faethm Pty Ltd – Written evidence (AIC0141)

About this submission:

Faethm Pty Ltd is an Australia-based R&D firm that has recently launched an AI analytics product, “Tandem”, to help governments and companies around the world to understand the impact of AI, Robotics and Automation. Current Faethm contributions to policy and strategy include discussions and work with:

1. Senator the Hon. Arthur Sinodinos, Minister for Department of Industry, Innovation & Science, Australian Federal Government. Use of Tandem to inform Australian industry, employment and investment policy.
2. Hon. Chris Bowen MP, Shadow Treasurer, Australian Federal Government. Use of Tandem to inform Australian Labor Party’s economic policy.
3. Hon. Ed Husic MP, Shadow Minister for the Digital Economy & Future of Work, Australian Federal Government. Use of Tandem to inform Australian Labor Party’s economic policy. Co-authorship of media article.
4. Mr Peter Varghese, ex Secretary of the Department of Foreign Affairs and Trade, appointed by Prime Minister Malcolm Turnbull to lead the Australia-India 2030 Economic Strategy. Use of Tandem to inform the impact of AI in Asian countries for economic policy purposes.
5. World Economic Forum, San Francisco. Use of Tandem by WEF Centre for the 4th Industrial Revolution, to equip WEF with insight and data about workforce automation.
6. UK Office of National Statistics. Meeting scheduled in October with Director General; Director of Digital Services; Chief Data Architect; Director of Methods, Data & Research; Head of Infrastructure & Architecture, to discuss how Tandem might provide information about AI’s impact on UK industry, communities, jobs and the supply and demand of skills.
7. Companies in Australia, US, France including Google, LinkedIn, Moody’s, BNP Paribas, BlueScope Steel, Mater Hospitals. Meetings with UK companies begin in October 2017.

Faethm is led by Michael Priddis, formerly a Partner at The Boston Consulting Group and Asia Managing Director of BCG’s technology innovation practice, Digital Ventures. Michael is a UK citizen, who has led work with companies and governments globally on technology strategy and the implications of AI. Faethm’s team includes executives with backgrounds at BCG, SAP, Accenture and Macquarie Bank. A PhD- and MBA-qualified data scientist leads Faethm’s data and AI development work. A leading academic, a senior BCG data scientist and a Silicon Valley-based AI venture capitalist sit on Faethm’s advisory board.

Tandem: <https://tandem.ai/> Faethm: <http://www.faethm.ai/>

[The Guardian](#) interview with Michael Priddis, “Computers are good at the jobs we find hard, and bad at the jobs we find easy”

[Australian Financial Review](#) article, co-written with Hon Ed Husic, MP, Shadow

Minister for the Digital Economy and Future of Work, "*Australia unprepared for automation of its workforce*"

Submission response:

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

3.1. There are many research reports from organisations globally about the impact of AI on jobs, work and the future supply and demand of skills. Many of these are bleak and are framed at the macro national economic or industry level. To date, no report has been able to show the effects of AI on a specific location, company, business unit, job or individual, nor has it been able to show the effects of different types of technology, or show the future effects over time, or link financial information such as salaries or costs to modeling. Finally, while many reports describe the positive effects of AI on freeing workers from mundane or dangerous work, and some describe some of the skills future work may demand, no report has been able to translate these trends to the modeling of future demand for jobs and employees in an actionable way.

3.2. Faethm's analytics platform, Tandem, uses Machine Learning and a community of experts to do exactly this, then delivers the results to users via an online Saas product.

3.3. Use of Tandem allows governments and industry to forecast and drill into the effects of AI across all of the above parameters, and to understand how to plan industry, investment, education, welfare and communications response across geographies.

3.4. We advocate Government leadership, using tools like Tandem, to inform both public sector and industry responses, and to allow citizens to understand the actions they might take to ensure their continued participation in the workforce, as AI changes the nature of work and the supply and demand for skills significantly and quickly.

3.5. Additionally, Tandem is currently a product for governments and companies, but a consumer version will launch in 2018, to equip individuals around the world with insight about how work will change, and how they might access learning and development services to better prepare for the future of work. We advocate government support for this type of "bottom-up" and citizen-centric approach, to complement the "top-down" policy and strategy already underway in many countries.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

- 4.1. At the international level, Tandem delivers a forecast of the effects of AI on any company of any size, in any industry or location or country. Faethm’s accompanying Future Workforce Index shows the effects of AI by industry at a national economic level, and can be drilled into to see the effects by gender, location, salary range, etc, depending on the availability of input data. The Future Workforce Index is currently built for Australia, and is being built for US, UK, India, and Bangladesh. It will be extended over time to include countries across Asia, as part of Faethm’s work on foreign affairs and economic policy for the Australian Government, and could be extended to include any country, with or without that country’s involvement.
- 4.2. We believe that tools like Tandem and the Future Workforce Index may be of particular interest to the UK Government to ensure that the UK benefits from AI, and is not disadvantaged relative to other countries, especially given Brexit and the UK’s need to establish new trading relationships and international economic collaborations with countries around the world. Insight about the effects of AI on other countries, on their industries, workforces, economic implications and social effects, will allow the UK Government to develop more effective economic policies. It will also allow more informed policy and targeted investment in UK industry to better enable competition and successful outcomes with companies overseas. For example, the effects of AI and robotics-driven automation of garment manufacturing in Asia will impact UK retailers and logistics companies, while the effects of AI on business process outsourcing in India will affect UK service companies. Equally, understanding the future demand for AI by country and being able to show the commercial nature of this demand will enable UK companies seeking to export to those countries. Finally, being able to model and forecast the commercial, economic and social value of different technologies that are driven by AI, such as Fixed Robotics, Mobile Robotics, Advanced Materials, Social AI and Process AI, will aid UK domestic investment in the technologies of most value.
- 4.3. At the domestic level, the issue of most concern to us is wealth distribution problems caused by those with repetitive cognitive, information acquisition / processing, low cognitive level or manual skills finding their work automated at scale and soon, and unable to acquire new skills fast enough to remain in work, while those delivering new work via AI finding their economic circumstances benefiting massively. This outcome represents a major challenge to many countries, and is not yet clear ideas like Universal Basic Income or a “robot tax” will address them in a meaningful or sustainable way. We remain sceptical of both for several reasons, and instead take the view that new work will arise, as industries and customer demands evolve, if workers can evolve too. The challenge will be for every country to provide the measures needed to adjust work for its citizens in order to evolve too, and not find that slow or inadequate responses limit their country’s ability to benefit in a globalized

world. A consideration for the UK Government in this regard is not only the automation of low skill work done by UK citizens, but automation of low skill work currently done by overseas nationals resident in the UK, and what the UK Government's response and financial investment will need to be for these people, should this happen.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?
 - 6.1. As described above, Tandem and the Future Workforce Index forecasts the effects of AI on any industry, in any country, over time. These effects need not be either positive or negative; what is critical is how industries transition.
 - 6.2. For example, AI and robotics offers huge cost benefits to Supply Chain companies through the automation of driving, warehousing, stock management, logistics planning and information flow processes and jobs, but brings significant and expensive challenges of major scale and operating model changes to most of the industry, due to the relatively unsophisticated and manual processing of many companies in this market. As such, while large companies in this industry with greater financial resources and expertise will benefit as they deploy AI and related technologies, workers in all companies and smaller firms with fewer resources will suffer from the automation of work and reducing margins as bigger players out-compete on costs.
 - 6.3. However, as AI and Robotics automate current work and delivers cost benefits, the advent of these technologies offers considerable opportunities for new value propositions, business models, work and new companies, driven by mass use of drones and robotics, micro-distribution centres, vertical warehousing and advanced analytics that anticipate and ship for ever-more accurately modeled and changing consumer demand, all of which will increasingly rely on AI.
 - 6.4. Considerable numbers of new jobs, either in current companies or in new businesses that arise to deliver and support these new technologies, offer a target state for current workers and companies to transition to. If this transition can be informed and guided with good data and insight.
 - 6.5. A key point here is that, while Supply Chain companies currently lag other industries' adoption of new technologies, their scaled physical and real-world not intangible services prevent erosion of market share by internet-based overseas competitors or much smaller start-up attackers, if they can respond in time.
 - 6.6. Similar scenarios could be described in Healthcare, Financial Services, Education, Retail, Manufacturing - we see beneficiaries in all industries. The challenge for government is to equip all industries in all locations to evolve, so that economies transition effectively. The

opportunity is for countries to effect this transition in a way that increases the competitiveness of their industries and workers on a global level. We believe that the joined-up, integrative and holistic policy development that this requires is a major task that requires a shared level of data and insight, which is why Tandem is being deployed.

6 September 2017

Family Law Partners – Written evidence (AIC0089)

Submission of Evidence to House of Lords’ Select Committee on Artificial Intelligence

Author: Alan Larkin

Background: Solicitor, collaborative lawyer & Head of Innovation and Technology at Family Law Partners

On behalf of: Family Law Partners (UK) Ltd

Date: 5th September 2017

Preamble

This submission will address two questions posed by the Committee in the areas of:

- Impact on society (Qs 3 & 4)
- Industry (Qs 6 & 7)

This submission will be based on the practical experience of Family Law Partners (FLP): a niche family law practice founded in 2011 with a specific focus on promoting non-adversarial alternatives to the court process. FLP sought to leverage technology to assist its clients dealing with relationship break up. The technology we were looking for did not exist so we began to develop our own: commencing with a ‘low-tech’ legal triage platform and then encompassing data capture and analytics to machine learning and AI. This submission seeks to demonstrate:

- How small and medium sized enterprises (SMEs) can access, develop and deploy AI
- How the deployment may mitigate policy deficiency in legal services provision
- The power of data to deliver client-centric services
- The obstacles to SMEs resourcing AI development
- How inter-sector collaboration and focussed governmental assistance can help

Impact on society

Q3. How can the general public best be prepared for more widespread use of artificial intelligence?

1. By experiencing AI in their everyday lives particularly if AI offers a solution to, or at least the mitigation of, a problem. The problem can be framed for

the purposes of this submission as the difficulties encountered by individuals seeking affordable legal advice from suitably trained and qualified family law specialists (normally solicitors, barristers and chartered legal executives). This is an access to justice challenge. In the specific context of family law advice, the challenge has been exacerbated by a policy decision reflected in LASPO: the withdrawal, with some exceptions, of public funding (Legal Aid) for individuals experiencing family relationship breakdown.⁴⁵⁹ The private sector is unable to replace the gaps left by LASPO but has innovated to make privately paid legal advice more affordable. Pro bono endeavours have also increased but cannot match the resources withdrawn by the State. There are constraints to which various solutions - unbundling⁴⁶⁰, fixed fees, free initial consultations - can be made affordable over the period to time normally required to resolve legal issues, especially when faced with the complexity of the law, the constraints of regulation, and the growth of litigants in person (LIPs) creating delays in the court process and struggling with legal process overall.

2. In 2011, we began work on an online 'guided pathway' that enabled prospective family law clients to complete a legal triage questionnaire free of charge. This guided pathway of over 1,000 conditional logic questions prepared clients for an initial consultation with a solicitor. The triage platform is in effect a 'virtual' family law solicitor.⁴⁶¹ The prototype was launched in July 2014 and continues in use today with over 300 clients having used the platform. Clients no longer need to be charged for us to conduct a fact-find on their matters (a vital exercise as the law can only be applied to the unique facts of each client's circumstances which are often complex). The results of the fact-find are automatically emailed to a solicitor so that a first consultation can immediately address the options and remedies available to a client.
3. This guided pathway has delivered benefits to both clients and lawyers within FLP although there will not be sufficient space here to fully explore this in any depth. But by way of observation upon the best interests of the public we would comment as follows. Our development of technology is informed by our professional background and regulatory framework. Furthermore, our specific focus on forms of dispute resolution that attempt to keep our clients out of the court process means that the technology we have built has by default, part of this professional DNA, which simply put means that we always seek to act in the best interests of our clients subject to proper observance of the rule of law and the interests of justice. The guided pathway

⁴⁵⁹ The Legal Aid, Sentencing and Punishment of Offenders Act 2012 (LASPO)

⁴⁶⁰ Unbundling allows for a series of discrete, separately costed steps in a legal process to be carried out for an individual without the creation of a full solicitor-client retainer.

⁴⁶¹ See definition of guided pathways by Roger Smith OBE: "[Digital Delivery of Legal Services to People on Low Incomes](#)", Chapter 2.

denotes a subtle shift in the power of the relationship which presently resides with the legal professional. The guided pathway attempts to inform and empower the client about certain family law processes and issues before a first legal consultation. A better informed client, aware that there are viable alternatives to the court process, such as mediation, mediation support, unbundling, collaborative law model, early neutral evaluation and family law arbitration, is going to be more assertive and questioning if their chosen adviser seems fixated on an expensive court process.⁴⁶² As family law legal specialists we regard it as our ethical obligation to create technology solutions that mirror best practice, the highest professional obligations and the steady shifting of power and control over the direction of a legal matter from our hands to those of our clients.

4. The prototype of the guided pathway was the inspiration for [Siaro](#) which was developed as a separate platform offering a guided pathway to be accessible to other family lawyers outside of FLP. Development of the Siaro platform to facilitate unbundling and pathways to appropriate dispute resolution models for family law clients is presently suspended as the development costs could not be sustained. Applications for grants or tech incubator schemes were unsuccessful. We will endeavour to galvanise the development again as soon as finances allow not least because of the potential the model offers for anonymised data interrogation for the improvement of outcomes (for which see further below) and for AI enhancements to reduce legal costs to the public.
5. As regards preparing the public for AI, most of the public already encounter AI in their daily lives via subscriptions to AI enabled Software as a Service (SaaS) platforms such as Spotify, Netflix and Amazon. The GDPR⁴⁶³ will protect certain rights of the public as regards the management of the data submitted to and used by such platforms. We would question how many commercial firms are aware of their obligations as data controllers under the directive but the GDPR will provide strong protection for the public and the heavy fines for breach will provide a compelling incentive for firms to take data protection seriously.

⁴⁶² See Solicitors' Regulation Authority, Code of Conduct: O(1.12): "*Clients* are in a position to make informed decisions about the services they need, how their matter will be handled and the options available to them;

⁴⁶³ General Data Protection Regulation (GDPR) which will apply from 25 May 2018.

Q4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

6. We have constrained our submission to address family law in the legal services sector. The adoption of AI in legal services has been restricted, in the main, to the large 'Magic Circle' firms who have the resources to explore use cases for their clients. These clients are themselves, large corporate entities. There are obvious benefits to be gained from reduced lawyers' fees gained through the time saving abilities of machine learning software. Why pay a large commercial law firm for the hundreds of hours it would take to go through thousands of legal documents when AI can carry out the same task in a significantly shorter time? Or, for a telecommunications corporate with cross-jurisdictional issues, why not expand the in-house legal team and equip it with AI enabled technology to deliver time and costs-savings initiatives? In short, our view is that large, well- resourced firms or large in-house legal teams will be able to explore and develop AI with little difficulty if they have the will to do so. This will benefit corporate and commercial activity for large or global law firms.
7. The question we would pose is: "When will the benefits of AI more directly benefit the individual consumers of legal services?" Is it acceptable for private individuals in pressing need of legal advice but constrained by affordability to wait patiently for the digital crumbs of innovative AI to fall from the high table of the Magic Circle firms to the level of the high street? Our view is that our clients should not wait. Family law clients are facing the stress of life-changing events for themselves and their children and we would argue that their need to access affordable legal advice is as great, if not greater, than for corporate entities. If technology, including AI, can offer these efficiencies and savings and make legal services available to private clients of modest means then we should make such development a priority even if that means juggling the limited resources available to an SME.
8. Despite the cost of software development and the considerable cost of 'soft-time' that must be committed to such projects here are opportunities for SMEs to explore innovation in technology and in AI in particular. FLP began a Knowledge Transfer Partnership (KTP) with the University of Brighton in March 2017, led by Senior Lecturer Dr John Kingston and supported by Dr Andrew Montgomery both from the School of Computing Engineering and Mathematics (CEM), using artificial intelligence to develop novel models of family law provision, including the use of automation. Funding was awarded from Innovate UK. The result is that a software systems engineer has been embedded in the family law team at FLP to help us develop AI applications relevant to our specific sector. The placement will last for two years. A potential commercial application arising from this partnership has already emerged. We would submit that the KTP programme is an effective way to

promote collaborative inter-disciplinary working between business and the academic sector.

Industry

Q6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Q7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

9. We have addressed the legal services sector in particular and the challenges of funding for SMEs that may wish to develop their own AI applications if their particular area of expertise has not yet seen much innovation. We have specific experience of data analytics. We have already referenced the guided pathway that became Siaro. After some four years of use, the guided pathway has produced a rich source of data which we realised we could explore to improve client journeys and outcomes. We anonymised the data and uploaded it to IBM's Watson platform which allows the interrogation of structured and semi-structured data for helpful correlations. We explored the ability to arrive at more accurate estimates of cost and case duration which we are obliged, by the Solicitors' Regulation Authority, to provide at, or immediately after, a first client consultation. This endeavour is topical, following a report from the Competition and Markets Authority (CMA) in December 2016 which called for more price transparency on legal costs for individual consumers and small businesses.⁴⁶⁴ Our ambition, in the interrogation of the data held, goes far beyond the likely mandatory information legal firms will be required by the SRA to produce. We are interested in identifying the precise drivers of cost and case duration. Is it, for instance, the number or age of the children of a marriage, the level of conflict in a relationship, or the period of separation that has elapsed before legal advice is sought? Or is it an interplay of the multitude of other factors we capture in our guided pathway? The IBM Watson platform has enabled us to see the correlations. Our ambition now is to build a predictive model for costs and case duration that allows our clients to make informed choices on their choice of dispute resolution model. FLP is fortunate to have the data to interrogate. Data collection will not be available to most legal firms, even those in the Magic Circle. We can use the lessons gleaned from the anonymised data to hopefully make an informed contribution to any consultation on mandatory information reporting initiated by our regulator.
10. It should not be lost on the Committee that an SME of our size could not possibly conceive of developing its own, sector-specific, AI without the

⁴⁶⁴ Competition and Markets Authority: Report into Legal Services, December 2016

support and funding liberated by Innovate UK through its KTP programme. Cost is not the only barrier of course. There is much that could be said about the structural and cultural nature of the partnership model still prevalent in many legal services firms that inhibits the investment and innovation required to pursue software development. That is a different discussion. We would submit that for those legal firms with SME status who have the focus and drive to enhance their services through AI development, the KTP programme is a significant enabler.

11. The Committee has invited participants to define AI. We prefer to think of AI as 'Augmented Intelligence' rather than 'Artificial Intelligence'. Family law clients prefer to see their advisers in the flesh. They want access to humans who can offer empathy, guidance and understanding: what some call soft skills or EI (emotional intelligence). In the case of lawyers, and family law specialists in particular, the EI should be complementing the years of professional training and the highest standards of professional integrity overseen by a vigilant regulator. Couched in these terms, we could say that access to such human interaction – a face to face delivery – is the gold standard. If we accept that is the standard we should adopt then we should explore how the highly trained family lawyer possessed of EI should access augmented intelligence, to reduce risk, automate otherwise laborious, expensive processes and make it easier for private clients of modest means to access this gold standard. The Committee will note that we have avoided burdening this submission with too much technical language. We could have explained the different types of AI likely to be deployed in our KTP with the University of Brighton. But much like our clients and indeed our regulator, we are much more concerned with the quality of the outcomes rather than the means of delivering them.

5 September 2017

Dr Jerry Fishenden – Written evidence (AIC0028)

Submission to the House of Lords Select Committee on Artificial Intelligence

Dr Jerry Fishenden FIET CITP FBCS FRSA, Visiting Professor University of Surrey. This submission is made in a personal capacity. 25th August 2017.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1 In domains related primarily to perception and cognition – such as facial and speech recognition, or detection of unusual patterns of trading activity indicative of fraud – “AI” techniques have proven the ability to learn effectively and to bring significant benefits (as Amazon’s Alexa is demonstrating, along with other areas such as autonomous vehicles). Many of these advances have been assisted by improved processing power alongside the focus within well-defined domains – evidenced in declining error rates within those domains – rather than earlier efforts which attempted to tackle “learning” across a much broader stage.

1.2 We are still nowhere near the “Turing test” type of “AI” that the public expect – i.e. a system that can think, create and generalise as well as a human being. Generalised self-learning systems remain confined to the realms of science fiction: we still await unsupervised learning systems able to generalise successfully and convincingly across broad domains.

1.3 Unfortunately “AI” has become an often largely meaningless label applied to software asserted to be adaptive and “self-learning” to the task at hand rather than being pre-coded. All software is written by humans and contains all the biases, mistakes, errors, conceits and failings of its creators, either by accident or intent. It will also be impacted by the nature of the training data utilised, which may be incomplete, partial or biased in some way. The boundary between “AI” and other software is fluid and ill-defined. **The Committee should therefore consider not only those programming techniques labelled “AI” (however that might be arbitrarily defined), but software in general where the ethics of the impact of that software need to be clearly understood.**

2. Is the current level of excitement which surrounds artificial intelligence warranted?

2.1 Partially, in some specific domains (as per para 1.1), although “AI” has

experienced regular periods of hype ever since 1955 when John McCarthy minted the phrase. It also has a tendency to self-promotion and overstatement – such as the grandly named “neural networks” which come nowhere close to the true neural composition and functioning of the human brain. Some successful data-related statistical work labelled “AI” may well draw upon techniques such as bayesian inference and heuristic analysis. Many so-called “AI” techniques are often applied for pointless activities – such as trying to reverse engineer consumer behaviour to guess what they may be interested in in order to serve up adverts that irritate them. A simpler and less computationally expensive method would simply be to ask consumers what they are interested in. Other applications demonstrate more toxic outcomes, such as the so-called “surge pricing” of taxis during terrorist events in London and Sydney.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence? In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

3.1 We lack a good level of understanding of computers and software in society in general, of which the “AI” issue is only a subset. “AI” systems are already in widespread use at companies such as Apple, Microsoft, Google and Amazon, in video games, and for facial recognition for security and understanding and improving flow management at airports. Most such systems complement and assist rather than replacing humans. They help improve productivity and also let humans focus on their strengths – creativity, interactivity, etc. – leaving lower value, high scale repetitive volume work to computers which can handle large data sets and analytics more efficiently.

3.2 We need a concerted effort to improve the public understanding of technology in general, through a well-resourced “public awareness and understanding of technology” campaign with a good reach into all sections of society. We also urgently need to have more expertise embedded within Whitehall and Westminster, and on company boards. There is a pressing need for far better technical advice inside government – one option is to consider creating the role of “Chief Technical Advisors” to help advise Parliament, politicians, Ministers, Permanent Secretaries and government departments. The current Chief Scientific Advisor community has few technologists or technological expertise amongst its membership despite technology’s critical role in the modernisation of government and our public services – leaving departments unable to understand or leverage technology advances, or to recognise their economic and societal impacts early enough to develop relevant policy or regulatory responses.

3.3 Please see paragraph 7.1 for wider considerations concerning security, privacy and data ownership.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

4.1 Most significant financial benefits at present appear to be accruing to the companies deploying claimed “AI” software. However, areas such as facial recognition have brought benefits in areas such as security, including airport security, albeit with associated controversy around their use, including aspects such as bias, discrimination and false positives. Other areas of application include anti-fraud systems in financial areas as well as an increasing role in healthcare in terms of assisting with analysing and identifying issues of clinical interest in complex medical data.

4.2 Many of these uses appear to be running well ahead of any associated ethical, regulatory and legislative regime, something likely to undermine public trust in such systems or, worse, create profound negative impacts within specific communities, with consequential negative reactions towards government and certain business sectors. Organisations most able to benefit from “AI” are those that own the bulk of the data in a particular sector, industry, or space. “AI” systems are primarily limited by the training data available to the individual system. This may well be only a subset of the overall data – as for example with health data, where the NHS has access to only a subset (that of sick patients) and not the wider set of health-related data held outside NHS systems (such as in wearable devices, gym equipment, smartphones, etc.)

4.3 A notable negative aspect is that “AI” can further entrench and narrow people into their own echo chamber. Online recommendations for example that *“People like you also bought products like this”* risk narrowing people’s experience. Alternative techniques that ensure people are exposed to wider choices and options – *“People like you never read articles like this”* for example – and early work on things such as the [Syzygy Surfer](#) need to be explored. In addition, we need to see a far better understanding of these issues to ensure appropriate public trust in these systems, assured where appropriate through regulatory, contractual and legal means.

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

5.1 Yes. There needs to be a comprehensive, transparent programme for better informing and engaging the public in the discussion of “AI” and software in general. This needs to begin soon, rather than letting misapprehension,

misunderstandings and falsehoods take root that will be difficult to displace – as they have been with other complex topics such as GM foods and fracking. There needs to be a better understanding that the current state of “AI” remains relatively primitive and most successful in narrow, specific domains. It is highly desirable for all “AI” research, learnings and developments to be openly published for wider review and understanding, moving away from an environment where academic publication happens behind paywalled obscure journals, and where commercial companies make absurd marketing claims unsupported by meaningful peer-reviewed evidence or proof in the public domain. As per para 3.2, government also needs far better technical capabilities, such as could be provided by experienced technical advisors embedded within policy making.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? *In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.*

6.1 Historically industries and jobs that have been supplanted by automation have been blue collar jobs (such as those replaced by robotics). Machine learning and “AI” applications will however begin to supplant white collar, professional workers. Specific domains of professional work that are primarily the application of research and facts, and identification and analysis of patterns, are some of the areas most likely to be impacted quickly. Less obvious, but equally likely and perhaps more critical, is the automation of component tasks – for example, material selection decisions in architecture, that might previously have been made by a civil engineer. While this will disrupt many existing jobs, as with other technological changes it may not necessarily decrease the number of workers employed, but will enable them to focus on increasingly value-added roles rather than the lower level work that is becoming better suited to computers.

7. How can the data-based monopolies of some large corporations, and the ‘winner- takes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

7.1 Government should be leading by example in the secure, consensus-based use of data and the establishment of general principles to be applied to the ethical use of software (including “AI”). Issues where governments can help establish principles and standards include:

7.1.1 user consent: engaging and educating users to ensure their consensual participation and understanding including of the data they are revealing, what is done with that data and how they may (or may not) be able to provide or revoke

consent

7.1.2 legal context: consideration of the legal context and how far e.g. the Digital Economy Act (2017), Data Protection Act and GDPR apply to machine learning, internet of things etc. or how they may need to be updated to keep pace with changing technology

7.1.3 economic: the impact that “AI” and other software are likely to have at both micro- and macro-economic levels in the UK, including on the potential future configuration of UK public services as the IoT and embedded health sensors etc. become more ubiquitous

7.1.4 access and control: establishing a trust framework across these many systems and humans’ relationships with them, one that spans anonymisation, pseudonymisation and strong identity proofing

7.1.5 data quality: data needs to be of sufficient accuracy and veracity to ensure that resulting decisions are coherent. This is a complex area – consider for example just one field, health, where the quality of many patient health records is unknown. Before building analytics and machine learning on top of such unknown data quality, users should be provided with access to their data to ensure their records are accurate. Many environments need to have precursor mechanisms in place to assess and improve data quality – including assessing data sets for inherent bias. Software-enabled or supported decisions are likely to amplify the bias of poor or inaccurate data and lead to inappropriate or potentially damaging outcomes. Consider commercial organisations such as Facebook building large international biometrics databases and related tracking systems based on users tagging faces in photos: this assumes that people are accurately tagging and not intentionally or accidentally misidentifying data. There is an inadequate focus on data quality, and the pyramids of assumption, analysis and decisions being built on what may actually be worthless or badly distorted data

7.1.6 data de-identification and anonymity: known problems already exist with anonymising personal data successfully and this has become an increasingly significant and complex issue. De-identification is not the same as anonymisation. More research is needed in this area to look how far e.g. attribute confirmation / exchange or techniques such as differential privacy might be more viable (or more appropriate in specific contexts) rather than providing access to raw data, whilst still enabling beneficial applications of machine learning

7.1.7 data access: ensuring appropriate control mechanisms for data (public and private / personal) accessed by such systems including appropriate protections (security / privacy / audit / accountability / protective monitoring etc.) are in place

7.1.8 data veracity / integrity: how do we know that data being used by such systems can be trusted? How do we know all data have been released from the systems when attempting to regulate or ensure they are compliant with e.g. laws of non-discrimination?

7.1.9 metadata: improving the understanding of the role this will play and how much use it is likely to in reality (as opposed to academic theory) e.g. see Cory

Doctorow's 2001 thoughts on metadata's true value

7.1.10 code jurisdiction: whilst some code may run within the UK (in particular systems, devices, or sensors) much will be operating in the cloud, or in private data centres or interacting with other systems scattered across the planet. There is a need to clarify how UK and non-UK systems will operate particularly in terms of whether they meet standards required (e.g. not exhibiting biased, illegal or discriminatory behaviour, or being compromised by hostile actors)

7.1.11 resilience: as many goods and services become ever more reliant upon this new generation of interconnected systems, the potential resilience to failure (accidental or malicious) will become an issue. Research is required into the potential interactions and vulnerabilities and risks of emergent systems of systems. It is also likely that all such systems will (a) need to be readily isolated from their environments should they behave in an undesirable way or be compromised by hackers, malware, etc. (b) be remotely patchable, requiring secure mechanisms to do this since, as with SCADA (Supervisory Control And Data Acquisition) systems, remote management facilities themselves present a potential vector for security compromise

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? *In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.*

8.1 If the right approach is not taken, the downside of this emergent generation of systems is that they will be discriminatory, wrong, biased, unaccountable, manipulative, and create significant security, privacy, legal and trust issues. However, if well applied the upside is that they will help support better policy-making, health care, education and transport etc., through responsive and more efficient systems. **These are ethical issues that apply to all software and should not be limited to so-called "AI" software alone.** However, government appears ill-equipped to develop appropriate ethical frameworks – see for example issues with its data science ethical framework. Note also the further detail provided in paragraph 7.1 above.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

9.1 Consistent standards of security, privacy and software engineering together with transparency about the decisions such systems are making is required. Systems and the decisions they make or enable must be able to demonstrate when challenged that they behave in unbiased, non-discriminatory and non-

invasive ways and are making applicable, acceptable and legal determinations. The data that they rely upon and on which they have constructed their models needs to be trusted, accurate and verifiable. Any exceptions to this need to be identified quickly and early so that appropriate remedial and corrective action can be taken.

9.2 The most viable option is probably to assume a “black box” approach and therefore adopt a model that requires certain data to be made openly available by systems to enable analyses of observable external behaviours, including longitudinal analysis over time. This could involve making sufficient data available via open interfaces (APIs) so that the external characteristics of systems and services can be inspected / analysed and held to account. Consideration needs to be given as to how open such interfaces and data would need to be: genuinely open (to all) or open to specialists? This will likely vary by subject domain. There is also the issue of how to ensure data is being fully released, and how to assure the integrity of that released data (i.e. that it has not been modified in some way to game the system and make it appear to be unbiased when in practice it is). After all, data released from a system might not be the same as the data held within the system. Appropriate issues of liability and insurance should be considerations here to help encourage the right behaviours.

9.3 There is also the issue of where boundaries are drawn – technical, legal, accountability etc. – in what will often be a complex ecosystem of interacting components using both “AI” and non-“AI” software and hence likely to exhibit sometimes unpredictable emergent behaviour. As the European Commission’s working document on the internet of things points out, any such interdependency “gives rise to a number of questions, such as:

- *Who is responsible for guaranteeing the safety of a product?*
- *Who is responsible for ensuring safety on an ongoing basis?*
- *How should liabilities be allocated in the event that the technology behaves in an unsafe way, causing damage?” [p.22]*

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

10.1 It would be a mistake to try to isolate “artificial intelligence” or “machine learning” or any other name given to self-learning software from any other software-based processes. Agreeing on what is “AI” as opposed to other software based techniques will prove a frustrating activity and distract from the core issue – which is how to ensure that software performs as expected. Current regulators appear incapable of performing this function and ill-equipped to regulate and hold to account the software and its creators that increasingly run almost every

business (consider for example the Volkswagen emissions issue and problems with the operation of Uber's software).

10.2 Society needs to be assured that software is not discriminatory, or rigging the system or otherwise failing to operate in a trustworthy way – whether someone decides to label such software “AI” is immaterial (particularly given so much of it is “black boxed”). It matters little to a citizen whether an inappropriate, wrong or discriminatory decision is made by one type of software system or another: what is needed is trust in decision-making / assisting software, particularly those operating in domains such as health. Along with the public awareness of technology and its social, political and economic implications outlined in paragraph 5, the role of education also needs to be considered – not just the narrow focus on coding and computer science, but on the creative and social sciences to ensure not just jobs for future generations, but also to equip them better emotionally to understand, navigate and deal with the future.

10.3 The recent Digital Economy Act (2017) is notable for entirely missing the need to include devices, sensors etc. within its definitions (it assumes existing administrative information systems, citizens and officials), missing most of what digital government and digital society is rapidly becoming. What is needed are highly precise bills / regulations / codes of practice that ensure compliance: technically agnostic law is often inadequate, hence why we have RIPA, the IP Bill etc. which incorporate tech within them. A similar approach is needed for trust in software. To do this we need genuinely expert groups, working in the open (see e.g. <https://www.gov.uk/design-principles#tenth>), both to get the best possible outcome as well as building public trust in what is being developed / proposed.

10.4 The underlying issue is the behaviour of digital devices / systems and digital machine ecosystems, not just their learning characteristics (which are a subset of the problem space). **So the policy issue to be addressed is a broader “trust in machine behaviour”**. Such machines will include devices and sensors around us in the growing internet of things (IoT), including software running in hardware and firmware.

10.5 Requiring minimal standards for software engineering / quality could be one potential approach (e.g. ISO9126, application of e.g. CISQ, and inputs from the National Cyber Security Centre, NCSC). Consideration is required as to whether there are some minimal trustworthy computing requirements that could be developed / used / stipulated, particularly for use in more sensitive domains (health especially).

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

11.1 Some good work has been done, but has often led to the arbitrary distinction of “AI” from other software techniques – when in fact the same principles of trust and transparency are required regardless of the nature of the software utilised. This is particularly true given the very grey lines around what is “true AI” versus “AI washing” etc. **If only “AI” software is regulated, some industries, companies, suppliers etc. may decide to stop labelling their systems “AI” to avoid such regulation – another disadvantage of such an arbitrary distinction.**

11.2 Some relevant work to be considered includes:

- US Federal Government Automated Vehicles Policy September 2016
- Royal Society machine learning dinner (28 July 2016) and their ongoing work at Machine Learning
- European Commission Staff Working Document: Advancing the Internet of Things in Europe

25 August 2017

Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams – Written evidence (AIC0029)

CALL FOR EVIDENCE – ARTIFICIAL INTELLIGENCE

PRELIMINARY STATEMENTS

Expertise from The University of Edinburgh

1. Since the mid-1960s, The University of Edinburgh has been actively researching Artificial Intelligence (hereafter AI), with the contributions of an estimated 2000 person years of academic and research staff, and an estimated 1000 PhD and 2000 MSc students, all specialising in AI topics. The School of Informatics (with colleagues from other Schools) is the largest academic AI group in Europe. The School's wide-ranging and extensive effort has been applied to advancing general AI methods and a range of AI specialisms, including natural language processing, machine learning, robotics and computer vision, planning, knowledge representation and reasoning. The results of this research have also been widely applied in many sectors. While developing their own research agenda, there is still considerable interaction between the specialisms, in part driven by shared underpinning technologies (e.g. probabilistic modelling, data science methods, deep 'neural' networks, machine learning, knowledge representation and reasoning).
2. *Although there has already been a huge investment by the UK in AI research, training and technology transfer, we collectively believe that the development of AI still remains an exciting long-term endeavour, and AI will be one of the defining technologies of the future.*

Definition of Artificial Intelligence

3. *AI, as currently realised, is not what is seen on television or in the cinema. It is a pervasive and powerful technology, but it is not yet a general purpose technology. It is currently deployed as a performance enhancing component in a range of highly specialist applications. These can be reasonably straightforward tasks (e.g. simple precision agriculture, car driver emergency braking, camera face detection, smart-phone predictive text, speech transcription and generation, smart search, enhanced household appliances). Or they can be more complex decision-making processes such as natural language understanding, machine translation, self-driving cars, IBM's Watson and personal assistants like Apple's Siri.*

Some are clever applications, but their abilities do not go beyond their narrow domain.⁴⁶⁵ The methods underpinning these applications are not new and magic technologies. Instead, they are cleverly engineered collections of advanced computer algorithms, including optimisation, search, knowledge representation, data mining, machine learning, sense data analysis, as well as deep networks and robotics. The result is that AI is better defined by its applications than by its underpinning technologies.

4. *The "intelligence" of an application needs to be distinguished from the "autonomy" of an application.* The former gives the application enhanced capabilities; the latter gives the application independent decision making and the ability to act. AI applications have varying degrees of intelligence. Few (to date) have autonomy, and that autonomy is usually closely constrained and formulaic, e.g. in an autonomous vehicle, a business-to-business purchasing agent, a stock-trading agent. Conversely, there are many autonomous and semi-autonomous systems such as self-guided missiles, nuclear power plant emergency shutdown systems, and aircraft autopilots that do not exhibit the kinds of intelligence AI is concerned with.

5. *Many of the House of Lords consultation questions are not simply AI questions.* Issues of privacy, liability, economic displacement, monopoly, transparency, governance, licensing are relevant to the broader modern economic ecology; AI is only one component. For example, a largely automated factory invokes many of the same issues, but need not be based on substantial AI elements.

Where the real dangers of AI lie

6. *The real and current dangers of AI do not lie in superhuman AI, irrespective of what one sees in the cinema or hears in the media.*⁴⁶⁶ There have been some major AI successes where performance is close to or exceeds human skill, e.g. autonomous vehicles, hand written character recognition, speech transcription, speech generation, partial machine translation, partial text understanding, and selected areas of medical diagnosis. Each of these very narrow competences is the product of 30-50 years of research by hundreds, if not thousands, of scientists and engineers. The human mind is claimed to be the most complex mechanism known to humans - replicating its sophisticated and general capabilities is far beyond current capabilities.

7. *Nonetheless, there are genuine dangers arising from widespread use of AI.*

8. The ability to compute at high speed and large-scale means that significant disasters can arise from automated reasoning errors or inadequate

⁴⁶⁵ A. Darwiche, Human-Level Intelligence or Animal-Like Abilities, CACM, in review.

⁴⁶⁶ A. Bundy, Smart Machines Are Not a Threat to Humanity, CACM, 2017.

understanding of the fragility of complex interconnected systems before humans can intervene (e.g. the 2010 stock market “Flash Crash” exacerbated by automated High Frequency Trading algorithms). Similar vulnerabilities arise elsewhere because it is hard to predict all consequences of complex interacting systems. This is especially the case when the algorithms within each system are commercial or military secrets, as was the case with the stock trading systems involved in the flash crash.

9. Economies can decline by being out-competed given the additional leverage of AI in algorithmic decision making and automation (e.g. business-to-business sourcing of cheapest materials, large-data analysis of trends and other business information, agent-based modelling of economic scenarios, flexible factory automation). This is in addition to the risks of losing UK income due to the improved commercial prospects of competing products whose performance is enhanced by AI methods (e.g. predictive text in mobile telephones).

10. Social unrest can increase dramatically due to reduced opportunities for meaningful employment, as a consequence of automated manufacturing capability (and the concentration of wealth to those who can afford to invest in it), and of the displacements of middle-level skilled labour replaced by automated service systems (e.g. travel agents, sales executives).

11. Smart, but indiscriminant weapons. They will have limited targeting mechanisms, and will be prone to incorrect decisions. Automated object recognition algorithms have advanced greatly but performance typically varies from 10-90%, depending on the types of objects and number of categories. Even a 1% false positive rate could have a devastating impact on civilian populations. For example, consider the consequences arising already from the wide-spread use of land-mines, which are passive weapons mainly affecting non-combatants. With a little AI, they can actively respond to or even seek targets, based on heat-signatures and movement. They are vulnerable to being hacked or left behind, possibly damaged, after a conflict, causing unintended damage long after the original conflict.

12. Social unrest could increase dramatically due to the speed of change and innovation. AI methods could be adopted widely and at large scale due to their economic advantages. Consider the impact of “smart-phones” on different social generations. Human society has not experienced this rate nor type of change previously.

13. Social problems could arise due to widespread ignorance about the capabilities of AI enhanced systems. People are familiar with the inadequacies of speech understanding systems (e.g. the humorous Burnistown “11” elevator video⁴⁶⁷). But most people are unaware of the AI enhancements in the products

⁴⁶⁷ <https://www.youtube.com/watch?v=sAzUvnUeuU>

and systems that they use. Instead, the understanding of the non-specialist is largely shaped by the mainstream media (newspaper humour and scare stories, high-profile “end of the world” statements, exciting but unrealistic movies). The consequences of the lack of understanding are unrealistic fears and unrealistic expectations.

Specific Recommendations

14. We note that these recommendations arise in the context of the discussion on AI, but are, in fact, also relevant to non-AI technologies, such as data collection, storage and analysis, data science, advanced manufacturing, video surveillance, and social media.

15. **Economics and Employment:** Because digital objects can be easily replicated and distributed, popular products can easily lead to concentrated wealth-generation by a few dominant market actors, as seen by the rise of e.g. Microsoft, Google, Amazon, eBay, and Facebook. Thus, innovative models of wealth and benefit distribution are needed. Bill Gates suggested to tax robots⁴⁶⁸, but this could be extended to more general AI systems. In response to the displaced human labour, we advocate an increase in training and employment opportunities in human-based services (e.g. healthcare, ageing, teaching, social care, activities, tourism). It’s particularly important for people who would previously have been employed in low-skill jobs that will cease to exist. Any ‘living-wage’ would need to be set at a level that enables people to participate in these services and the general economy. Substantial economics research will be needed to develop models for how an economy with decreasing amounts of human labour might work and how the benefits of the society will be distributed. Global equality issues will become more urgent (and consequences, such as migration).

16. **Safety:** Most AI is embedded in products and systems, which are already largely regulated and subject to liability legislation. It is therefore not obvious that widespread new legislation is needed. Systems with embedded AI should be covered under standard recall and fault recourse mechanisms. Manufacturers should demonstrate due diligence, as with any other product. There are existing models of risk and methods for standards and their verification. These need to be enhanced, but not necessarily replaced. Additional legislation may be needed: 1) to provide a framework for requiring satisfaction of specified standards as part of the licensing for deployment of critical AI systems, 2) for situations where multiple independent AI components are integrated into a larger system (either on a single device or across a network), and 3) to address the issue of computer speeds, wherein actions happen at a time scale far faster than humans can

⁴⁶⁸ https://en.wikipedia.org/wiki/Robot_tax

respond or intervene. Enhanced developments in cybersecurity are needed to make AI apps safer and less hackable. A particular worry is the use of embedded apps by companies with limited experience with software development and computer security (e.g. car and household appliance manufacturers).

17. Privacy and Use of Personal Data: Because of the potential for widespread collection and automated collation of personal data, and their subsequent use in automated decision-making systems (e.g. insurance pricing, social benefit determination), it is likely that additional legislation will be needed to govern activity around discovery, access, deletion and correction of personal data. For example, one should have access and control over to the data collected by an “always-on” personal digital assistant. Another issue concerns what information can be uploaded for corporate analysis from these personal digital assistants. The provenance of training data should be explicit, and should be “fair”, e.g. representative of the variety of the human population recorded in the dataset.⁴⁶⁹ Legal and regulatory systems need to be enhanced (esp. considering the widespread “illegal” use of software “cookies” in the EU).⁴⁷⁰

18. Education and Public Awareness: People need to understand broadly the capabilities and limitations of AI enhanced systems and products, and how these capabilities are expected to advance with time. They should also be familiar with their rights and risks. Introduction could occur at school, but given the changes that will occur post-schooling, some web-based public information mechanism would be essential. Royal Institution lectures⁴⁷¹ are useful, but appeal to a narrow audience. Broader dissemination and engagement mechanisms are needed, particularly since the effects are likely to affect less skilled labour harder and earlier.

RESPONSE TO SPECIFIC QUESTIONS

19. Current and future state of AI

Current state: narrowly specialised AI applications are becoming pervasive, e.g. auto-correcting and predictive phone app text. However, there is no clear boundary between an AI-based application and other well engineered computing applications. Most AI enhancements are emerging as a convergence of 30-40 years of academic research (autonomous vehicles, natural language understanding, automated translation), large datasets providing examples of many variations of the recorded phenomenon (customer preferences, automotive

⁴⁶⁹ Royal Society Machine Learning Report 2017.

⁴⁷⁰ e.g. <https://www.theguardian.com/technology/2015/mar/31/facebook-tracks-all-visitors-breaching-eu-law-report>

⁴⁷¹ Prof. Chris Bishop, October 2016 Royal Institution “Discourse” on Artificial Intelligence

faults), improved machine learning and data mining techniques, and cheap desktop supercomputing.

Future developments: lots of increasingly sophisticated, embedded, special purpose applications will provide improvements to personal efficiency and informedness. There will be gradually improving general question answering systems. and widespread medical discovery and diagnosis systems. Increasingly capable and general object recognition systems and reliable and commercially feasible 2-legged mobile robots will follow. These developments are likely to accelerate the competition and gains of major companies and national entities with the resources to invest in research, development and deployment of AI systems. There is likely to be a thriving ecosystem of small AI players. These developments are also likely to accelerate the concentration of wealth in these companies and countries, leading to increased social problems, including migration pressures.

*Human level general intelligence*⁴⁷² AI is much further in the future.

20. Is excitement warranted?

Yes and no. There are increasing numbers of special purpose, incrementally useful applications. These will keep increasing. Although AI has seen many “hype cycles” and reassessments (and there are likely to be more), there have also been real gains, to the point that elements of AI technology are present in almost anything involving a computer.

21. Preparing the public and their understanding

As noted above, there should be relevant and regularly updated exposure at school level, and public awareness media for all ages. There could be training courses for “application advisers”, much as there are financial advisers at present. They could advise on which apps to use, how to connect and use them, and how to stay safe. This could be a commercial skill, but with training supported and encouraged by government.

22. Who will gain most/least?

Under current economic models, it seems likely that the big winners will be the organisations that have the resources to invest in AI technologies (national, military, or commercial). As a software technology, AI applications are essentially infinitely replicable. There could eventually be a small set of apps competing in most product areas (e.g. current software for most things other than smart-phone apps). The producers of these apps will become very wealthy because of the ease of production and distribution (e.g. sale of Microsoft software, small

⁴⁷² Which itself is not well defined nor understood.

transaction or usage license fees). In terms of the public good, everyone is likely to benefit from products and services with improved and more personalised services. Improved transport, energy distribution, manufacture and agriculture could reduce production costs. Improved medical diagnosis would benefit all. The reduced need for many types of semi- skilled human capital and the training cost and pre-requisites for high-skilled labour are likely to lead to an increasing pool of underemployed and low wage people.

23. Data Monopolies

Large personal datasets can be collected by any AI-based service, e.g. most data science- based services. But even non-AI based web services will collect large amounts of personal data, so issues concerning data monopolies are not just AI issues. Central concerns include: whether data can be sold, data security, provenance, and different levels of access and detail.

24. Ethical Issues

The core issues are safety, liability, and fairness. We have a concern for the potential for near- instantaneous disasters (e.g. like the stock-trading disaster), their scale, and responsibility for when they do occur. We need standards and legal liabilities for AI enhanced products, but based on the product itself, rather than on any particular aspect of the AI. The fair use of data is not strictly an AI issue: privacy, consent, diversity and the impact on democracy are issues that arise in the context of general big data, cybersecurity, and social media.

25. Transparency of AI

It would be ideal if an AI system could provide an intelligible justification for its reasoning. However, except for the simplest of rule-based systems, this is rarely possible. Logic and proof-based systems can reason based on hundreds or thousands of steps. Probability based systems generally use hypothesised causal relations with probabilities learned from collating possibly only a few, or possibly millions of instances. The recently developed deep learning methods tend to out-perform other methods, but their decision processes are numerical, and are generally completely unintelligible. *What seems more feasible is to only licence critical AI systems that satisfy a set of standardised tests, irrespective of the mechanism used by the AI component. Equally, one could question whether most human decision making is transparent and highly accurate.*

26. Role of government

Any legislation that affects the deployment of AI systems will need to be agreed internationally, otherwise the UK acting alone risks leaving itself at an economic disadvantage. The UK is well placed to further invent, develop, and exploit AI methods; any legislation should ensure that this supportive environment continues. Existing technical, economic, and social legislative mechanisms are

Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams – Written evidence (AIC0029)

adequate for the moment to cover most AI as well as other computer-based areas, such general software liability, databases and privacy, and cybersecurity.

27. We appreciate being consulted on this issue and hope that our statement provides a helpful contribution. We welcome the opportunity to continue the discussion if desired.

Prof. Robert Fisher FIAPR FBMVA and colleagues: Prof. Alan Bundy FRS FRSE FREng FACM, Prof. Simon King FIEEE, Prof. David Robertson, Dr. Michael Rovatsos, Prof. Austin Tate FREng FRSE FAAAI, Prof. Chris Williams FRSS

25 August 2017

Dr Malcolm Fisk – Written evidence (AIC0012)

1. My Expertise:

I am responding in an individual capacity you will wish to note the following:

*In his **De Montfort University** capacity Dr Malcolm Fisk leads the European Commission funded **PROGRESSIVE** project - see www.progressivestandards.org that is addressing 'Standards around ICT for Active and Healthy Ageing'. This project focuses on key issues that relate to smart homes, telehealth, co-creation and interoperability.*

*As Director of the **Telehealth Quality Group** (TQG) Malcolm is actively engaged in supporting the development of telehealth services according to appropriate service paradigms (see www.telehealth.global). At the heart of this work are quality benchmarks for telehealth (relating to a wide range of domains and very much from a service user / consumer perspective). This includes the development and promotion of a well-respected International Code of Practice for Telehealth Services.*

Malcolm's other roles include being

- a) expert advisor for **ANEC**: The European Consumer Voice on Standardisation. Representing the consumer interest he is a participant in three European CEN Committees - relating to standards for (a) health ; (b) quality of care for older people; and (c) social alarms.*
- b) member of a Quality Standards Advisory Committee for **NICE**, the National Institute for Health and Care Excellence.*
- c) expert Advisor to the **European Commission** Coordination Hub for Open Robotics.*

*Previously Malcolm was appointed by **Welsh Government** to Chair the National Partnership Forum for Older People and subsequently to provide expert advice on for a relating to addressing poverty and inequality; and the housing and related support needs for older people.*

2. Definition of AI:

As I do not directly work in the field of AI I am not positioned to give a technical definition. A more rounded definition that allows for the human / societal might look like something like this ...

AI is the 'intelligence embedded in devices and systems that that they can operate in accordance with programmed instructions, external stimuli and, importantly, automated learning.'

3. Pace of Technological Change:

Q1:

I am not well positioned to comment

Q2:

The current level of 'excitement' is not warranted given the dangers and challenges that require to be recognised and responded to. Some of the latter are noted below.

4. Impact on Society:

Q3:

I am expert in the field of digital health and am, therefore, concerned that AI should be harnessed for individual and societal health and well-being. Linked with this is the necessity of social change in areas that your 'call' notes for matters such as privacy, cyber-security and data ownership.

There are linked 'fields'. These include eHealth, telehealth, telecare, smart homes, robots and the 'Internet of Things'. In all of these - consideration needs to be given not just to the nature of the 'interface' between the technologies and the individual users (who may, in many cases, be vulnerable and lack physical or cognitive abilities); but also to the very purpose and meanings around the tasks that the technologies are intended to perform.

In preparing for the more 'widespread' use of AI it is necessary for

- a) manufacturers and suppliers of products and services to shed what are sometimes patronising and ageist views of older people. Such views result from old social norms and constructs that view older age as associated with dependency (consider e.g. the 'retirement' age; 'dependency' ratios; and the specific, generally institutional, types of separate and often segregated accommodation such as sheltered housing and residential care).
 - 1.
- b) certain safeguards to be put in place because of the vulnerability of many consumers / users. This reflects the fact that many older people may not have well-developed digital skills and, as disproportionate users of health and care services, are particularly at risk to cyber-criminals, e.g. for identity theft, because of the 'richness' of health data (in terms of the personal information).

And in consideration of the way that AI can change through 'learning' (e.g. from technology usage, activity monitoring, etc.) there is a massive consideration that relates to (i) the nature of (often usually inappropriately lengthy and complex terms and conditions) that apply to goods and services purchased; and (ii) the

way in which users / consumers can give consent for the same (this is even without consideration being given to the knowledge, and any consents required, of carers).

You ask of data ownership! There is only one answer ... and that is that personal (and certainly health) data is the property of the individual to whom it relates. In the health context this data is 'entrusted' to the service provider (and this includes public sector services through e.g. the NHS and local authorities). You may wish to note the perspective of the Telehealth Quality Group (www.telehealth.global) on this is as follows (from the International Code of Practice for Telehealth Services) where the clause (D1 below) sets out the position. Many other provisions within the Code (e.g. such as those relating to cyber-security) are relevant too.

Protecting Personal Information

Requirement:

Services shall maintain current policies and procedures for the management and protection of personal information. An exception applies (see below).

Applicability:

Applicable to all services.

Guidance:

These policies and procedures shall ensure that services operate in a manner that is fully in accordance with country, state, province or region specific legislative or regulatory requirements. The policies and procedures shall give attention to the transfer of personal information over publicly accessible networks and the manner in which such information is accessed - whether via fixed or portable devices. Specific procedures for the protection of personal information shall be included.

Policies and procedures shall ensure that the manner of storage, management and sharing of personal information normally carries the explicit and informed consent of users and carers. It follows that such consent shall be renewed prior to any proposed change in arrangements for the transfer or storage of personal information. Staff shall be precluded from storing data on their personal technologies/equipment except in authorised circumstances (such as when on-site and alternatives are not possible).

In this context, services shall demonstrate an understanding that such personal information is owned by the users and carers themselves. It is, therefore, entrusted, with their consent, by users and carers to the service for the contracted period and can only be retained by a service provider in certain circumstances.

Exceptionally, but only when authorised by law, the need for explicit consent may be overridden.

Policies relating to the management and protection of personal information shall be posted on the website. They shall be dated and reviewed annually. Services shall be guided by the principles set out in ISO 27001.

ISO/TS 13131:

This clause together with D5 covers Clauses 11.2, 14.1, 14.2 and 14.4 of the ISO/TS 13131 Health Informatics - Quality Planning Guidelines for Telehealth Services.

Q4:

Who is gaining the most (and least) from AI? In any developments around technologies there are winners and losers. There is, however, an absolute imperative for manufacturers, suppliers, researchers, product developers (etc.) to pursue designs that are inclusive and safeguard privacy. Their 'touchstones' must include the tenets of 'Responsible Research and Innovation', RRI) as promoted by the European Union – and where various initiatives are being promulgated by my University (De Montfort, Leicester); and 'privacy by design' (the concept developed in a joint initiative by the Dutch Data Protection Authority and the Privacy Commissioner of Ontario).

In sum we need to point to the need for simple, usable, intuitive interfaces; and functionality that allows for people's control and empowerment ... and an 'on-off' switch.

5. Public Perception:

Q5:

The best ways to improve public perception depend on building trust. Our reference point here is the most recent Eurobarometer survey (2017) which notes (from fieldwork undertaken in March), for respondents, that 74% (in the UK) would 'make more use of digital technologies if there were more widespread tools to improve reputation and trust'. An indication of the extent of distrust is the fact that 40% of respondents, over the past three years, were 'less likely to give personal information on websites' and that 36% expressed their readiness to 'pay more for better security and privacy features when buying IT products'.

Linked to this is the effect of publicity relating e.g. to data breaches (NHS institutions are particularly high profile) and straightforward reporting as provided by Which? who (by using ethical hackers) found about half of a selection of commonplace household items were insecure. I'd like to know what the position is regarding health technologies – including those that I am most familiar with in the worlds of telecare and telehealth. My wariness on this matter directly relates to the emphasis given to cybersecurity in the TQG International Code of Practice for Telehealth Services.

Without trust (and in the context of continuing publicity of data breaches) there will be no improvement in public perception.

6. Industry:

Q6:

I am not well positioned to comment – except to say that there is a real potential in the 'world' of digital health ... subject to the provisos noted above.

Q7:

I am extremely concerned about the position of the mega-corporations and their position with regard to the holding of personal data. The European Commission initiated GDPR may help with this (e.g. with regard to the 'right to be forgotten') – but can it be enforced to the extent desired? But bringing things closer to the consumer we need, I suggest, some simple (generic) terms and conditions ... that relate to our personal data *and which would apply whether we are sending off for a part for the washing machine; ordering a new passport; or we are accessing (online, potentially via a video-link) a telehealth service in which our personal data is held and exchanged.*

7. Ethics:

Q8:

I have noted the issues of privacy and consent in response to Q3.

Q9:

The 'black box' issue is of extreme concern. With knowledge and intelligence confined to a tiny few (software engineers, etc.) there will be few (any?) 'touchstones' by which users / consumers can determine the 'trustworthiness' of the technologies, etc. that they wish to access or maybe provided for them. It follows that in this context one of the tenets of RRI (noted above) regarding 'openness' and 'transparency' takes on greater value. But in this rapidly advancing commercial world that we live in (where businesses may be predisposed to sacrifice their integrity on the altar of profit ... in order to give a fast return to investors) it is difficult to see how this might be achieved.

8. Role of Government:

Q10:

Regulation of AI is definitely required. This should build on the GDPR and be undertaken in collaboration with our partners in the EU.

9. Learning from Others:

Q11.

Others are probably better placed than I to advise on this. But I would suggest that consideration of the issues around AI would wisely give attention to its role in health and the (often linked) position of vulnerable people. With such matters in mind (around e.g. data privacy and security) a broader, more inclusive and more generally applicable approach might emerge.

With regard to grappling with the issue of terms and conditions and, maybe, having a simple set of generic conditions that could apply in different contexts (where different levels of personal data might need to be provided, stored, exchanged) I believe there has been some work undertaken in Sweden. It would be a good idea to consult with ANEC and/or Consumers International on this matter.

14 August 2017

Five AI Ltd – Written evidence (AIC0128)

FiveAI was founded in September 2015 with the ambition to leverage research in artificial intelligence to build autonomous vehicle technology that can be deployed as a force for good in the UK's transport sector. Today it has teams in Cambridge, Bristol, Edinburgh and Oxford and is well on its way to building a system that meets the safety requirements of dense road topologies, traffic, pedestrians, cyclists and behaviors common in urban areas across the UK and the world.

FiveAI's highly experienced team scientists and engineers are building on research to implement the complex software brain of tomorrow's autonomous vehicles. This software must be robust in all potential mobility environments and situations, be aware of other road users likely actions before they happen and ensure that vehicles powered by our software drive safely and just as other road users would expect.

Given the novel research we are undertaking and our application of artificial intelligence techniques to deliver important industry and societal change, we offer the following response to the call for evidence.

The pace of technological change

- 1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?*
- 2. Is the current level of excitement which surrounds artificial intelligence warranted?*

Impact on society

- 3. How can the general public best be prepared for more widespread use of artificial intelligence?*
- 4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?*

The use of artificial intelligence to develop autonomous vehicle services will result in (i) recovering ~230 driving hours/year per commuter improving quality of life and/or additional productive use of that time valued at £3,100 per user (ii) a sharp fall in mobility expenses of £2,100 per driver (lower depreciation, insurance, fuel, parking and maintenance expenses as AVs are in productive use 20% of available hours, not 3% today) (iii) increased load factors per vehicle (to average 2.5) at no material user inconvenience, so cutting traffic flow on heavy urban roads by an estimated 37% in London and higher in smaller cities (iv) fast switch to EVs, eliminating CO₂, NO₂ and particulate emissions, CO₂ savings on commuting alone would amount to 499 ktonnes pa by 2025 (v) progressive

improvement in vehicle safety towards 10x that of human driving, reducing damage to property, injury and death (vi) safe and reliable on-demand mobility to young, elderly and disabled citizens, including school runs, social and leisure travel and (vii) release of parking in urban and business clusters, representing eventually up to 5% of land use.

However, replacing human drivers with software could also have a negative collateral social impact, not least on those made redundant by the technology and their dependents. Furthermore, it's likely that the software replacing humans globally will be delivered by only a few, mostly western, companies; thus further concentrating value to a limited number of shareholders in those firms. Our expectation is that societal good will outweigh societal bad, albeit the former has an indirect impact whilst the latter has a direct one.

Public perception

- 5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?*

Industry

- 6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?*

The global (and local) market opportunity for artificial intelligence enabling autonomous vehicle services is huge. In 2015 the world manufactured and consumed around 87 million cars. McKinsey is one firm that has estimated the growth in the number of autonomous shared cars shipped by 2030, which they put at 10 million units that year, out of total shipments of 115 million vehicles. That would imply a global fleet of around 25 million autonomous cars by that date, but autonomous cars work much harder than owned cars, so they estimate autonomous cars could account for 32% of all journeys by that time.

Five AI believes this prediction is conservative and the ramp will be much sooner and the numbers will be much bigger, more like 20 million autonomous unit sales by 2025 and a 40 million autonomous vehicle global fleet by that date. If correct, this global fleet could offer capacity of 200 million seats. Assuming an average seat utilization of 50%, and 12 billable journeys a day per utilized seat, that would equate to 1.2 billion billable journeys each day. Assuming journeys would be priced the same as a bus ride (£4) the service opportunity alone equates to a £1,400 billion annual total available market (TAM). Our modelling suggests an operator net margin of at least 14% is achievable at these fare levels, assuming appropriate levels of vehicle, insurance, cleaning, maintenance and energy input costs.

The supplies of those inputs also represent very significant global market opportunities themselves by 2025, including: supply of vehicles (£400 billion), sensors and computational hardware platforms (£200 billion), software licenses and on-going support (£370 billion), insurance (£50 billion), energy (£100

billion), cloud/app IT services (£40 billion) and support services, e.g. monitoring, maintenance, cleaning (£220 billion).

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

1. In order for autonomous vehicles to interact safely in urban environments they must be able to predict the likely reactions and behaviours of other traffic participants (e.g. cyclists, pedestrians, cars, lorries, buses, etc.). These behaviour types vary per participant type, road topology, time of day, weather, etc. Their geospatial differentiation could provide some margin to avoid global 'winner-takes-all' dominance, i.e. localities may support their own duopolies.
2. Artificial intelligence techniques such as inverse reinforcement learning make it possible to learn the required behaviours from vast numbers of examples. At present, only large vehicle OEMs and their Tier suppliers have access to sufficient scale of examples, gathered from the vast numbers of vehicles they've sold that are operating on the roads.
3. CCTV, traffic cameras sources, etc., if appropriately pseudonymised, could provide an alternative source of raw data from which can be extracted the object and behaviour classifications that need to be learnt. Given the high number of these cameras under operation by various Government agencies across the UK, these could be 'opened up' to provide an alternative to the data silos of automotive industry incumbents.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

The application of artificial intelligence in the field of autonomous vehicles will pose a range of ethical considerations. The following areas are indicative of the types of issues we will face but are by no means exhaustive.

Training Data

Assuming a reliance on artificial intelligence-based machine learning techniques for perception, there is potential for classification bias, e.g. unequal precision-recall performance across age, gender, physical ability and even race. The data used for training perception classifiers should be curated to avoid this and, where possible, classifiers should be transparent to inspection to verify non-biased performance.

Decision Making

Though best efforts will be made to measure scene confidence and situational risk to reduce occurrence of trolley-problem type dilemmas to an absolute

minimum, artificial intelligence-enabled autonomous vehicles may still, on rare occasions, encounter such decisions. Though it may be possible for systems to evaluate more evidence, quickly, than is typical for human decision makers when faced with such problems, e.g. the age, physical ability or gender of potential casualties, it would be prudent for the companies operating in this space to agree to treat all human life as bearing equal value and not to make any distinction based on internal / external location, financial relationship to the service provider, age, gender, physical or mental constitution. This would be in line with the German Federal Ministry of Transport and Digital Infrastructure Ethics Commission report on Automated and Connected Driving (2017).

9. *In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?*

4. One of the most exciting prospects that AI brings is that of fully autonomous vehicles. This will require the use of both black box learning and of transparent decision-making systems, in tandem.

5. **Explanation:**

6. Machine learning is the dominant component in modern AI systems. The most recent and most powerful wave of machine learning – Deep Neural Networks or DNNs – are black boxes that ingress input data and egress decisions, without any clearly interpretable intermediate states. However, as the recent Royal Society [1, ch 6] report on machine learning states:

7. *"not all machine learning methods use this approach, and alternative approaches can be more readily interpreted."*

Systems that have more intermediate structure in their decision making are intrinsically more capable of explaining their decisions but may be less accurate at making predictions [1, ch 6.2]:

8. *"Machine learning methods could be restricted to those that directly yield an interpretation [...] However [...] there may be important trade-offs between interpretability and accuracy."*

9. The business of FiveAI is autonomous driving, and transparency is of the essence. Yet such complex systems will fail – indeed it will be a major achievement to build and deliver systems which achieve human-level safety, never mind exceed it, in our urban environments. Building and maintaining public trust, explaining the cause of errors and building corrective action programs will be paramount to that trust being deserved. The challenge is particularly bracing when one considers that 99% accuracy is rarely achieved in a single machine learning module, and yet the assembly of such modules into a system may need to achieve an accuracy of up to 7 x 9s to be confident it is safe in a fully autonomous setting. Human interpretable intermediate states and full transparency in information engineering systems

will allow our systems to be auditable for safety. In particular, reasoning with probability will be needed to track, at every level of information processing, the nature and extent of risk. Such a probabilistic approach is often considered by authorities to be good practice [2] for autonomous systems.

10. The quest for higher accuracy at the expense of interpretability however means it will not be sensible to outlaw the use of black box systems. Indeed people already accept this with human experts, such as pilots or surgeons, who cannot fully explain their own skills, particularly skills at “lower” cognitive levels such as hand-eye coordination, and would be substantially hampered if required to do so. Exactly how to have the cake and eat it is the subject of ongoing research and development – for example systems under development in which two machine learning systems work in parallel [3], one optimised for accuracy and the other for transparency.

11.[1] *Machine learning: the power and promise of computers that learn by example*. Royal Society Report, 2017.

12.[2] *Concrete problems for autonomous vehicle safety*. McAllister et al., Proc. IJCAI Conference 2017. Paper

13.[3] *Why should I trust you?* Ribeiro et al., Proc. ACM (2016). Paper.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

In the artificial intelligence application field of autonomous vehicles, comprehensive access to hazard scenarios and test cases will be essential for delivering safer systems. The Government might consider a mechanism for stipulating that autonomous vehicle programs testing or deploying in the UK share the real world hazard events that they encounter so long as making such data widely available does not diminish the UK’s ability to build companies with first mover advantage in its key markets.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

We suggest that the German Federal Ministry of Transport and Digital Infrastructure Ethics Commission report on Automated and Connected Driving (2017) provides clear and rational guidance on the ethical approach that companies using artificial intelligence to develop autonomous vehicle technologies should adopt.

Five AI Ltd – Written evidence (AIC0128)

Please do not hesitate to contact us if we can be helpful in any way in your deliberations.

Stan Boland
CEO – FiveAI Ltd

6 September 2017

Foundation for Responsible Robotics – Written evidence (AIC0188)

Noel Sharkey, Emeritus Professor of AI and Robotics, University of Sheffield and I am speaking on behalf of the Foundation for Responsible Robotics (a not for profit) which I co-direct

1. In the last few years AI has emerged from the shadows and entered the commercial and industrial world at a fast and accelerating pace. This is largely due to three main factors

- (i) An incredible increase in machine memory and processor speed
- (ii) The availability of extremely large data sets (big data)
- (iii) Improvements in machine learning methods such as deep learning

2. AI has gone through a considerable transformation since the 1950s and currently learning from big data is the dominant force in the field. So instead of having to spend long hours hand coding tasks, a properly designed learning machine can zip through billions of lines of data and perform the task satisfactorily.

3. This is clearly advantageous to the very large number of startups who are exploiting this new wave of AI commercially. It is so much easier than having to work laboriously to specify the task and code it. However, it comes with a cost.

4. The main disadvantages are

(i) It is extremely difficult to check vast repositories of data for bias e.g. gender and race. And the bias can be subtle and not-explicit

(ii) a machine trained with such learning algorithm will pick up biases inherent in the data set and in some instances amplify them. These can be non-obvious such as picking up on likelihood of depression when selecting job applicants.

(iii) learning algorithms such as 'deep learning' are black boxes containing large matrices of numbers. This lack of transparency means that they are no labeled variables to tell what features (or higher order features) have been selected by the machine in the performance of its function.

5. Where this technology will lead us in the next 5, 10, 20 years will very much depend on public trust and on legal protection of the human rights of individuals whose lives are impacted negatively. And it will also be affected by the introduction of new data protection and privacy laws.

6. We can only speculate about the development of AI technology. When the development is on an exponential curve it is easy to speculate that it will

continue like that. However, it is just as likely, as has always happened in AI, that once all the low hanging fruit has been picked, severe limitations will be found and then the tech will plateau (whilst still being useful). There has always been too much crystal ball gazing in AI and frequently ambitions are converted into predictions.

Impact on society

7. There is a major problem with AI decision-making that requires urgent regulation. AI is being used in so many areas of our lives (and we can look to the US for what is coming) to make decisions. It is already being used to assess insurance premiums, to determine the best applicants for jobs, to decide on who should get bail, to help determine length of prison sentence and very many more examples. Massive evidence is amounting daily that demonstrates that these decisions are often very biased in terms of gender and race. This is mainly because the big data on which the AI systems are trained is historical and ossifies and amplifies our biased values.

8. This is then compounded by the learning machines themselves. The lack of transparency means that it is very difficult, if not impossible, to work out an explanation for the decisions being made. These are not labeled variables such as 'risk of pregnancy so don't employ'. We cannot find out from the matrices of numbers what is going on. This can create and perpetuate injustice and violate our human rights.

9. We are also seeing the rise of AI decision making in robotics in areas that include: the care of the elderly, child care, military targeting, policing, transport, delivery, sex, surveillance and autonomous transport. A future robot carer in the home of an elderly person could make decisions that erode their basic rights. Over the last 10 years in writings about ethical and societal dangers of the use of this type of robot, it has become clear that great care must be taken against the potential risks to many of our basic human rights:

- autonomy
- privacy
- freedom
- dignity
- human contact and communication
- right to life
- right to peaceful protest
- welfare
- security
- distributive justice
- wellbeing

10. Until now we have had good control of AI decision making in robotics compared to AI applications in other spheres as mentioned above. But it will not

be long before robot decision-making will follow suit in using large data to train the decision-making.

We must be very careful what decisions and control we cede to machines

11. This is not scare mongering about machines suddenly rising up and taking control. That seems highly unlikely given the limitations of machine intelligence. Machines do not have intentions or desires and it does not seem plausible that they will have any time soon. That acts a distraction from the real problem of misuse by humans whether deliberate or not.

12. The immediate concern is that by ceding decisions or control to machines, the humans start accepting their decisions as correct or better than their own and stop paying attention. And yet, as mentioned above, there is a growing body of evidence that the learning machine decision makers are inheriting many invisible biases among their correlations.

13. Urgent action is required

We need to act now to prevent the perpetuation of injustice by new regulation. It is important not to stifle innovation with regulation while at the same time protecting the public. At present any company can sell a decision algorithm to any other company prepared to buy it. There are no guarantees of unbiased performance. We need to specify clearly in law that

- (i) if any algorithm (including one on a robot) is shown to have an adverse impact on the lives of an individual person or a group of people, then that machine should be shut down until it is investigated.
- (ii) everyone should have a right to know why a decision has been made that impacts their life. This is difficult for black box algorithms at present and so they should not be used in this way until explanations can be derived from them
- (iii) large scale clinical type trials should be held to determine if a decision algorithm can be used fairly, without prejudice or bias.

Public Perception and trust

14. It makes good commercial sense to ensure public trust in AI and robotics or we may never see many of the positive benefits that can be offered. At present, public understanding of AI and robotics is largely based on science fiction from novels to Hollywood blockbusters. While these provide great entertainment they interfere with the reality of the technology. This leaves the public susceptible to many crank suggestions that are often amplified by the news media.

15. It is essential to ensure that public understanding keeps abreast of the technological developments in AI or there can be no informed public decisions or consent. A well informed public is needed to maintain public trust and lack of

public trust will inhibit developments and innovation. Such inhibition could come about either because of accidents or bad AI decision-making or simply by not meeting expectations.

16. A standard way of informing public perception is by means of talks and seminars from scientists and public demonstrations. However that has not been greatly successful as attendance numbers are limited and these are one-offs. To get real results we need to embed an understanding of AI and robotics through our education systems. Every child should have a number of small and large projects on the topics throughout their school career and thus understand what is going on as they reach adulthood.

6 September 2017

Professor John Fox – Written evidence (AIC0076)

AI evidence (House of Lords)

Professor John Fox

Definition of Artificial intelligence

1. I have adopted a definition for AI based on a benchmark for “autonomous cognitive agents” that several colleagues and I published in 2003 [3]. This is a traditional perspective based on a range of cognitive capabilities rather than a single function such as machine learning, though it includes learning as a necessary capability for general intelligence.

Main points

2. **Re questions 2 and 5:** Excitement about machine learning (“AI”) on the basis that it will produce a revolution in white collar services and professional practice is overblown, and appears more driven by corporate marketing and fashion than by evidence. In complex domains such as medicine talk of revolution is naïve and likely to be counterproductive.
3. **Re questions 1, 5 and 10:** Building general and human-level AIs (“AGI”) is scientifically challenging but a case can be made that it is already possible to build certain kinds of human-like and general purpose cognitive system. For reasons of social benefit, national competitiveness, societal risk and UK security this possibility should be investigated as a matter of urgency.

Author bio

I am a career cognitive scientist. My PhD was in experimental psychology (MRC Applied Psychology Unit, Cambridge, 1974) followed by a NATO/SRC postdoctoral fellowship with AI founders Allen Newell and Herbert Simon (Carnegie-Mellon University, Pittsburgh). After this came a stint with the MRC again where I worked on medical decision making and the use of computers to help make clinical decisions. Unfortunately the proposition that computers might make decisions as well as or perhaps better than doctors proved at that time to be controversial and in 1981 I was recruited by Sir Walter Bodmer, Research Director of the Imperial Cancer Research Fund (now Cancer Research UK). For the next 25 years I ran an interdisciplinary group working on theory and design of AI and its potential to help improve quality and safety of patient care.

I have published widely in cognitive science and AI, computer science and software engineering, as well as in the medical literature. I set up was editor of the *Knowledge Engineering Review* (Cambridge University Press) and have served on many editorial boards. With Subrata Das I co-authored *Artificial*

Intelligence in Hazardous Applications (AAAI and MIT Press 2000) which is believed to be still the only comprehensive discussion of the issues and options for designing and using AI in safety critical domains such as medicine. I have unexpectedly become a serial entrepreneur, having led several medical AI start-ups including Expertech Ltd (1986, acquired by Inference Corp), InferMed Ltd (1999, acquired by Elsevier) and Deontics Ltd (spun out from Oxford University and UCL/Royal Free Hospital in 2014). My current passion is OpenClinical CIC, a non-profit company whose goal is to empower non-technologists - like doctors and patients - to exploit AI to improve quality, safety and user experience in humanly appropriate and acceptable ways.

Context of this evidence

There are several debates about artificial intelligence currently raging in the public sphere and in academia.

- A leading issue is whether AI (specifically machine learning) will replace large numbers of professional people working in "narrow" or niche and routine applications in healthcare, law, finance and many other sectors (ref Susskind and Susskind).
 - There are also concerns that AI and "black box algorithms" are used in important personal and societal roles but people are unable to understand or control them (ref Pasquale).
 - AI research could even lead to the creation of "superintelligence" and the possibility that an existential crisis will arise if autonomous, superintelligent machines are free and able to act against our interests (ref Bostrom)
 - There is also a dispute within the AI community itself about what AI really is. The GOFAI (Good Old Fashioned AI) community considers NEWFAI (New Fangled AI based on machine learning) as inadequate for building machines with a wide repertoire of cognitive capabilities comparable to humans. A new term "Artificial General Intelligence" (AGI) has been coined to avoid confusion and a distinct research community is emerging.
4. **Re questions 3, 4 and 5:** These and other debates have captured the imaginations of the media and the public to an extraordinary extent. This has been partly driven by the extraordinary power of consumer devices to "understand" speech, "recognise" people and objects, "plan" and navigate journeys, "drive" autonomous vehicles and so forth. Claims of major capabilities in AI by major companies based on successes in game-playing have also raised expectations. There is a widespread expectation that all these narrow capabilities will in due course be integrated in big and culture-changing products, such as autonomous homes, factories and weapons, though it is not yet clear whether integration is tractable or how it might be done.

Artificial intelligence in medicine

Re questions 1 and 2: Learning not programming.

5. Traditional software development has become notorious for spiralling costs in design, implementation and maintenance, difficulties recruiting and retaining skilled developers, and the problems of legacy code, particularly in large-scale business and safety critical applications. It is not therefore surprising that managers who are responsible for procuring and supplying software products and services should be attracted by the idea that in future computer systems will “program” themselves: learning how to carry out tasks and progressively improving their performance through “experience” without the need for a human programmer or supervisor.
6. The truth appears to me to be rather different, in that it is only possible to build any software system once we know what the task is (we want to win at GO or give the best treatment to a cancer patient) and what the potential risks and costs of failure are (not just losing a game but killing a patient), so there must still be a great deal of sophisticated human work to be done before learning can even start. This is certainly the case in medicine; if there is good objective evidence that the economic benefits of learning technologies over traditional software development are as great as many say then achieving clarity about when this is the case and when not will be of great value.

Re questions 2 and 10: Evidence of success and the hype cycle

7. Healthcare is a major target for AI companies, small and large, and generates great interest among general and business readers. This encourages commercial companies to use their efforts to apply machine learning and AI to put out statements about their developments as though they have already achieved good results and successful adoption by health professionals. Media aggregators then take all these statements and uncritically republish them to give the sense that there is a true revolution going on without hard evidence that this is the case. The wide interest in medicine then encourages people to believe that if AI can have such a revolution in medicine then the potential must be equally great in other professional fields. Determining the true picture is vital and the parliamentary committee could play an important role by commissioning a dispassionate and objective study of the evidence for success in healthcare or otherwise.

Re question 1: Capability and versatility

8. As a scientist my primary interest is in the “big” end of AI, not in building physical robots but in how to design and engineer cognitive agents that are capable of carrying out a wide range of tasks in complex domains. Medicine has been an inspiring context in which to attempt this. It has led to a benchmark that identifies a set of necessary and possibly sufficient capabilities that a doctor (or any medical AGI) must be able to carry out if it is to do its job well, and the benchmark may not be limited to medical expertise. Medicine has also allowed us to show that GOFAI methods are capable and versatile, though it has long been known that they can be brittle in the face of the uncertainties characteristic of real world healthcare. NEWFAI methods have the opposite features – they can be robust in the face of statistical variability in the real world but, with respect to our benchmark at least, they have a limited cognitive repertoire.

Re questions 8 and 9: Black boxes and ethics

9. As with “big data” expectations about AI may already be moderating, and for at least one similar reason – the big algorithms that process, interpret and learn from thousands or millions of practical cases and use the results to advise us or even take autonomous action are opaque to human users. We may neither know nor particularly care what the precise ranking of businesses in the report that a search algorithm delivers when looking for an Italian restaurant in our locality, but (to take an example from an IBM Watson marketing video) I do care when Watson says that treatment A is 30% more likely to give me a good outcome than treatment B but my oncologist can’t tell me why, or how good the evidence is or reassure me that the pharmaceutical company that makes treatment A hasn’t “gamed” the algorithm to promote its products. In my experience of working in healthcare GOFAI and its subfield “knowledge engineering” make it far easier for people to understand what an AI is doing and why.
10. Pasquale’s *Black Box Society* identifies many personal and societal dangers of big data and closed algorithms. Neither “AI” nor “autonomous agents” appear in the index compiled around 2014 but I imagine Pasquale’s views on the growth of AI since his book was published, and prospects for effective regulation of black box AI would be illuminating should the committee seek them. At the very least a discussion of how to get the transparency and controllability of GOFAI alongside the benefits of NEWFAI should be promoted.

Re questions 1 and 2: Theory of general AI

11. My knowledge and understanding of AI’s implications are of course limited by the theoretical and applied problems I have worked on. I have been particularly informed and therefore probably biased by the idiosyncrasies,

needs and challenges of healthcare. However, medicine is arguably the most diverse and complex domain of professional practice in existence, and in my opinion there are few practical or theoretical problems that arise in other professional domains that are not also to be found in clinical practice. I am also concerned that in the present febrile debates influential claims about the revolutionary potential of AI are often made without evidence or technical argument and by relatively unqualified observers rather than AI practitioners; even practitioners can often be researchers with narrow technical interests and limited practical experience. While not wishing to state my own opinions in overconfident terms I think that lessons learned in applying AI to major problems in health and social care may be important for society in trying to understand what AI can do and what it can't, and how we should use it and how we shouldn't.

Design and engineering of AGIs

Re questions 1, 2 and 10: Scientific convergence

12. The question of whether and when we will be able to build human-level AIs and possibly even "superintelligences" is much subtler and deeper than predicting the short-term economic and employment impacts of machine learning. However, most practitioners believe that AGIs are still a considerable way off, and many think that there will never be a superintelligence that will out-perform people on a wide range of tasks. I don't agree with either of these positions and close by explaining why and offering a recommendation that I believe is relevant to UK policy and priorities and implications for our international competitiveness and security.
13. In 2014 Nick Bostrom, director of the Oxford Future of Humanity Institute published "Superintelligence: paths, dangers, strategies". Nick is neither a practitioner nor a researcher but he has a broad view of "existential risk" and his book restarted an important debate. In one chapter he speculates about several distinct sources of superintelligence and in another possible technical routes to human level AI (and from there to super AI). Although he suggests a number of paths to human-level AI he also reports that confidence the AI community's confidence in our ability to build a human level intelligence soon is low, being only 50/50 by the middle of this century and confidence does not become high (90%) until late in the century.
14. However there is an important absence from Prof. Bostrom's list; he does not make any comments on the possibility that the cognitive science research communities might achieve a theoretical breakthrough leading to an unexpected advance in our capability to engineer general purpose cognitive agents. Perhaps few of my colleagues would agree but I believe that there are reasons to take quite seriously the possibility of a significant, imminent and rapid theoretical advance.

15. Synergy between basic science and practical application development is often key to technological advances and this is no less true in cognitive science. My own group's focus on medicine forced us to develop technology that has capabilities over a wide range of clinical problems and settings, and to test theoretical ideas in the real world of medicine as well as in the lab. On our benchmarks we are may already be close to being able to routinely design cognitive agents that have breadth and depth of expertise that no single human clinician could match. As remarked I know most about AI in medicine but I think that the same could well be true in other fields of human expertise.
16. The challenges raised by medicine have been so great that we have also been forced to learn about and draw upon concepts and techniques from many different disciplines, while at the same time being forced to bring these ideas into a single design framework to achieve practicality and scalability from an engineering point of view. My involvement in many branches of cognitive science and neighbouring fields has led me to believe that there is an underlying convergence of thinking across many of them, including psychology and neuroscience, AI and computer science, decision theory and philosophy of mind, to name a few. Again I recognise the need for caution but I think this convergence might be heading towards an early breakthrough in our ability to routinely design and build certain kinds of human-like AGI.

Re questions 5, 6 and 8: The road ahead

17. Some of my colleagues might say this claim is preposterous and irresponsible and that AI's history of underestimating the complexity of human cognition and overstating its ability to deliver true machine intelligence has repeatedly damaged the field. Outside the research community there is less diffidence about making sensational claims, with major figures like Gates and Musk, our own Rees and Hawking and a large international celebritariat, saying that we ought to be worrying about the emergence of human-level intelligence (not just machine learning and autonomous weapons) and we ought to be worrying about it now.
18. I do not think that I am inclined to exaggeration but the likely economic and competitive advantages to the UK of having industrial strength in cognitive systems engineering, and the contrary risks of others having strength in the field while we do not makes the argument clear: the question of whether we could bring together current scientific knowledge from many different fields in order to build forms of AGI today, even if only limited ones, needs to be investigated urgently and rigorously. If, as I suspect, this is possible then the UK must capitalise on its strong cognitive science base and develop an equally strong AGI engineering capability in a sound ethical framework.

References

1. Bostrom N *Superintelligence*, OUP 2014
2. Fox J and Das S *Safe and Sound: Artificial Intelligence in Hazardous Applications*, AAAI & MIT Press, 2000.
3. Fox et al "Understanding intelligent agents: analysis and synthesis" *AI Communications* 2003
4. Fox J "Cognitive systems at the point of care: The CREDO program" *Journal of Biomedical Informatics*, 2017
5. Pasquale F *The Black box society* Harvard 2015
6. Susskind and Susskind *The future of the professions*, OUP 2015

5 September 2017

Laurence Freeman and Fabia Howard-Smith – Written evidence (AIC0147)

Laurence Freeman and Fabia Howard-Smith – Written evidence (AIC0147)

[Submission to be found under Fabia Howard-Smith](#)

Fujitsu – Written evidence (AIC0120)

Fujitsu welcomes the opportunity to respond to the House of Lords Select Committee on Artificial Intelligence. This submission sets out our thoughts in response to the Committee's questions. We have addressed those questions where we can offer a deep, genuine insight and would be happy to expand on this submission with a meeting with the Select Committee should be this of interest.

In order to avoid any confusion it is first worth a brief look at the how Artificial Intelligence (AI) is defined before addressing the questions posed by the Committee.

Defining Artificial Intelligence

A challenge with defining AI is that we are still not sure exactly what real, human intelligence is. A simple view would be to describe it as "the simulation of human intelligence by machines". AI relates to getting a computer to reason and to learn, and then to use this thinking as the basis to make decisions.

AI systems are excellent at pattern recognition; they can quickly spot anomalies and make predictions, often more consistently, more accurately and more reliably than humans. However, AI systems today are limited only to that. They use probability and logic to make their analysis, but lack the ability to comprehend or develop broad context as humans can. Such an ability, which could be called 'general intelligence', is still a long way off from current technologies and may never be realized at all.

Unlike most traditional computing structures, today's AI systems are not centred on massive, complex central processors. Instead, they are based on neural networks, modelled on the human brain, wherein large numbers of processing elements or nodes manage a mutual flow of information. AI is not a single, well defined entity, but incorporates many capabilities, models and methods. However, three elements in particular account for the huge acceleration and advances in AI of the past five years:

Machine Learning – a set of techniques – e.g. algorithms such as decision trees – that enable machines to learn from data, without being explicitly programmed for the task at hand.

Neural Networks – a computing model that arranges large numbers of processing nodes, from tens of thousands to millions, linked by an even larger number of connections, resembling arrangements of neurons and synapses in the human brain. The power of the system comes not from the individual nodes themselves, which carry out only simple tasks, but is derived from the layered architecture of the neural network as a whole, which becomes adept at recognizing complex patterns.

Deep Learning – a machine learning technique that exploits the architecture of a neural network with several layers, some of them possibly specialized for certain characteristics and patterns. One notable example is applying deep learning to recognize a picture of a cat. A typical neural network is six or seven layers deep; the most sophisticated now contain hundreds. The technique requires data – and lots of it – to work, but having been trained by looking at thousands or even millions of pictures, a neural network becomes very good at its task, better even than a human.

The real power is that the system only needs to learn once. Once learned, the system's knowledge – 'what does a cat look like?', 'what do normal data packets (as opposed to a security breach) look like?' or 'what does an unhappy customer look like?' – can be transferred to other applications, providing instant help in making decisions or recommending intervention.

It is also worth noting that we often bundle other technologies, such as robotics, into the same conversation as AI. That's because AI and robotics are such complementary technologies, with AI enabling automated decision-making and robotics enabling the decisions to be fed into physical actions. For instance, autonomous (self-driving) vehicles are the result of combining AI and robotics.

The Pace of Technological Change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Artificial Intelligence and automation are not new ideas. The concept of using neural networks to process data began in the 80s, and the manufacturing industry has been utilizing automation for a long time. What has arisen recently is an environment that makes AI a possibility, where modern speeds of computing, combined with the vast data availability, has enabled a huge surge in the application of AI and automation in the workplace.

The primary factors in the rapid growth in AI have been:

- Falling costs of computers and processors, enabling the construction of large neural networks.
- Massive data availability – providing more information that we can use to train neural networks.
- New techniques, architectures and algorithms enabling exploitation of these two developments.

We have an unprecedented computing power at our disposal, especially given special graphics processing units (GPU-s) that outperform normal CPU-s, via the FPGA technology (hardware accelerated algorithms) and special-purpose architectures. With them, we can use much “deeper” learning models, e.g. substantially more layers of neural networks.

Further, mathematical modelling and algorithmic advances, as well data availability – sourced from social media and the Internet of Things (IoT) applications – help the AI models learn faster and more accurately. In the next 10 years Artificial Intelligence will penetrate both private life (internet, cell phone based assistance, voice command systems, medical care, etc.) and business life (intelligent traffic systems, manufacturing, process automation, smart cities, smart energy use, etc.). Companies will utilise AI in Digital Transformation projects for a fundamental transformation of all parts of all value chains.

AI is not just the next wave of the digital evolution, it is an absolute necessity for dealing with the complexity and security challenges created by the recent networking revolutions i.e. the internet, mobile internet and IoT, which have done so much to transform the way in which we live and work. The amount of global investment in Artificial Intelligence and machine learning is sky-rocketing and the future of entire industries seems to hinge upon the success of AI technologies. For example, the predicted future of the automotive industry currently centres on the development of connected and autonomous vehicles.

The key to unlocking the potential of machine learning systems is the ability to feed them the massive amount of data they need to learn. Thanks both to the proliferation of IoT sensors and the huge volume of text and image data available online, we finally have enough data available to be able to train these AI systems to perform almost any task. This is important, because it is the training phase that is the most challenging for neural networks, particularly from the perspective of the processing power they require.

At Fujitsu, we’ve been working to harness AI to recognize patterns and images by turning data into images. By implementing what we call ‘imagification’, we have even applied AI to challenges that are typically not image-based. For example, we can interpret the movements received from a small accelerometer worn on a car driver’s wrist, by plotting them on a chart and training the system to differentiate between different types of movement. This could potentially be used by insurers to identify safe drivers.

It is clear that the days of ‘conscious AI’ are some way off. However, we are only just scratching the surface of potential implementations – in twenty years’ time when the technology has really matured, we expect it to have totally transformed every industry, from healthcare to retail to financial services.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

Artificial intelligence is hugely powerful – it may be the most powerful technology we have ever created. The true power of AI lies in its application. You can use AI to derive significant benefits from unsophisticated data sources, for example in monitoring CCTV feeds to manage traffic flow in cities, to spot suspicious individuals in public places, or to enable crowds at large events to disperse more efficiently, by guiding people to the most convenient exit or mode of public transportation.

Although they may not be immediately noticeable, the use of AI is already delivering improvements to our daily lives. Machine learning can enable a moving tractor to tell the difference, in real time, between a growing lettuce and a dandelion. A targeted application of weed killer can then give the lettuce more space to grow and deliver a more efficient crop yield. AI can also help determine the optimal time for harvesting – making more informed, intelligent decisions by studying weather patterns and historical information, as well as other data, such as current levels of supply and demand in local supermarkets.

Supply chain management is also significantly enhanced by the use of AI, monitoring inventories across entire production lines to make sure that supplies of essential components are never in danger of running out, and therefore avoiding expensive downtime. Human error and incapacity is wholly avoided by AI; machines never sleep, need a coffee break, lose count, or get distracted.

Fujitsu's human centric view is that AI will make humans more effective. Thanks to the assistance of AI, humans become able to work more efficiently, focusing on higher-value activities. This is exactly what the next wave of robots to arrive is helping us achieve. AI can help us make food production and supply chains so efficient that no food ever goes to waste. In medicine, AI can help doctors make rapid preliminary diagnoses, freeing up more time to address each patient's specific issues. Simple task completion by AI in customer service provides customer-facing staff more time for complex cases.

Impact on Society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

The rise of AI is part of the ongoing 'digital revolution'. Whilst much of this paper addresses the anticipated and unforeseen benefits of AI, as in the Industrial Revolution, the disruption caused by new technologies will necessitate as yet unpredictable changes in society, such as in education and the workplace, in order to adapt to the new paradigm.

A common perception of AI and automation is that it will cause a widespread loss of jobs. With many low skilled or process orientated jobs easily replaced by automation, this is understandable. Concerns around labour obsolescence are not new. However, since the Industrial Revolution, the labour market has always adapted and out of new technologies new jobs have been created, replacing those that have been lost. These jobs are often built on capitalising on new technologies, and that is where the focus must be placed to prepare society for AI to the benefit of the general public.

Rather than perceiving AI as something that threatens jobs, it is important to recognise the limitations of AI and understand how it can aid people in the workplace, rather than replace them. AI has certain strengths, notably the ability to process large amounts of data and spot patterns within them. However, AI cannot think about data as humans do. We will still need human input to make sense of those insights.

In this way, AI can aid people in their work rather than replacing them. It automates much of the manual work involved in sifting through datasets to find the significant information, from which point a human can take over and make a decision based on that information. For example, in the customer service industry, the addition of AI to service desks and call centres frees up staff from low-level, monotonous tasks, enabling them to concentrate instead on addressing more complex technical problems, or delivering better customer experience or care. In this example, employers may find that the ability to deliver an improved customer service experience with the same number of employees far outweighs the cost benefits of reducing the number of staff to maintain the previous lower level of service.

The growth of AI will also produce an entirely new market for jobs that we have yet to need. In the case of big data, the need to analyse new information produced a large number of jobs. The rise of AI will create a similar demand for new roles to analyse the insights being produced. Aside from people responsible for monitoring the output of automated AI systems, there will also be a greater need for programmers to support this new automated workplace and the rapid changes that will come with it. AI will require modellers to tell automated machines what assets they should draw upon and which patterns they should look for. These are just some of the new requirements that could create jobs as the automation revolution takes off – others may arise that we cannot presently predict.

By thinking of automation as a way to assist and enhance work, not replace employees, we can see a positive impact on society rather than focussing on damage caused. Crucially this means that the public will require new skills, retraining and redeploying to add value where AI and automation can't.

In order to prepare the general public for these changes that AI will bring to society, it is important for policy makers to ensure that the education system

properly prepares young people for a world where interacting with technology in a relatively advanced way will be an integral part of every job. Low skilled jobs are already being phased out in certain industries, such as in the retail sector where the replacement of cashiers is underway in favour of automated checkouts. Whilst the UK Government has made a start in improving technical education through the introduction of T-levels and a more general focus on developing effective vocational qualifications, the education system will need to be continually responsive to technological advancements to keep pace with changing demands in the labour market.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

With the rise of any new technology it is important to ensure that people are properly equipped with the ability to interact with it. For example, we still see a large proportion of older people who are not confident in using the internet, despite the many benefits that being online can have in alleviating issues that such people face, such as loneliness. Enabling these people to embrace new technologies can deliver significant personal social benefits, as well as potentially reduce costs in the care sector.

The development of AI has the potential to provide a similar scale of benefits to older, home bound people if they are properly used. Early stage home AIs, such as the recently launched Amazon Echo and Google Home, can be used in a range of functions to provide reminders, connect people with friends and family and perform simple tasks such as turning on the radio, lights or heating. Such capabilities are invaluable to those suffering from mobility or memory problems. As AI continues to advance and 'smart homes' become a reality, we could see a role for AI in the social care sector in providing such people with a level of self-autonomy. However, this is only possible if people are comfortable using this technology.

Charities such as Age UK help people get online and have a number of good programmes supporting this effort. If AI does begin to be used in the social care sector then it may also be worth considering funding training for interaction with AI devices through local care providers.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Traditionally it may be assumed that those sectors most likely to benefit from the development of AI and automation are those that are process or labour based. This is particularly the case in a more traditional definition of 'automation', for

example, robotics has revolutionised manufacturing processes to the point where products can be assembled by robots with minimal input from a human.

However, artificial intelligence stands to benefit almost all sectors of the economy over the next few decades. One area where we are seeing substantial benefits is in cyber security. In order to prevent cyber threats rather than react to them once they have occurred, we actively monitor the traffic of networks in real-time to identify potential risks before they can do any harm. We do this by taking a holistic view across the internet, monitoring all the traffic inside and outside of the systems we protect, 24 x 7.

No individual can do this, given the quantity of the data to be analysed. In fact, with the level of cyber attacks that are seen today, it is possible that we have reached the limit of what humans can achieve in cyber defence. Therefore, we use AI programmes to analyse the data with speed and accuracy. Our AI systems can learn what familiar patterns look like in a network and recognise normal traffic. Therefore, when it encounters data packets falling outside of these normal patterns, it immediately flags the anomalies. This machine learning is cumulative –it keeps on improving. Early in the training cycle, these systems raise a lot of false alarms, but over time they get better and better at identifying true threats.

However, there are clear areas where AI does not and perhaps is unlikely to ever be able to replace a human. While industries involving lots of manual work or methodical tasks lend themselves well to the use of automation through AI, those that require a more 'human' touch do not. For example, social care requires a level of empathy and compassion that we are unable to automate. It is highly unlikely that an AI will be developed with sufficient awareness to provide such emotions better than a human carer. Whilst the answer to question 5 mentioned that AI may be able to provide certain benefits to the social care sector, this does not extend to the human, compassionate care that many people need.

Similarly, AI could be used in the education sector to help create better courses or mark exams. However, the role of a teacher cannot be fully automated due to the different levels of interaction that students require and various learning styles that must be catered to in lessons. Whilst it is possible, as currently takes place, that independent learning can be carried out through interactive digital courses, the role of the teacher is unlikely to be fully replaced.

Besides the lack of genuine emotion, there is one fundamental reason why the current capabilities of AI cannot fully replace humans. Human beings have the ability to look at two completely unrelated things and make a connection even where there is nothing originally associating those two elements. AI cannot do this, it can only process data from the sources that it is given. Whilst it is feasible that such an ability may be possible in the future, this is likely to be far from now.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

It is important as the development of AI systems progresses, no AI is able to extend beyond human control or undertake tasks for which it is not meant. Currently AI capabilities are very limited and are mostly confined to automation tasks. However, as AIs become more sophisticated, potentially acquiring what we would call actual intelligence, limits must be in place to guard against unintended consequences.

Artificial Intelligence is generally recognised as having three calibres, or levels, that define its strength and capabilities. These are:

1. Artificial narrow intelligence - AI is in some areas superior to humans and limited to specific tasks
2. Artificial general intelligence – a single AI is equal to or superior in most areas to humans
3. Artificial superintelligence – AI is superior to humans in all areas

Artificial intelligence is currently firmly in the ANI spectrum, where AI systems are superior to humans in certain tasks. For example, AI systems have beaten the world champions of strategy games such as Go and chess, whilst other AI systems are able to analyse data and spot patterns far better than any human.

Even in this phase it is important that AI systems do not develop undesired biases and do not learn to act or make decisions contrary to our moral and ethical expectation. Even when these systems are quite simple and primitive, it is necessary to develop standards, rules and frameworks to prevent them being trained with material that introduces those undesired effects. This is a question of a societal debate and global agreement is likely to be necessary to secure this.

Should AI development progress to a more general intelligence, it will be necessary to have mechanisms in place that detect and prevent the use of non-certified systems. This in itself is likely to need AI systems to spot. Once again, global agreements must be in place to do this.

It is debatable whether we could create an AI that has superintelligence, however it is ethically questionable whether this should be allowed even if it were possible as it would by definition be beyond human control. An AI that extends beyond human capabilities would have the ability to create even more powerful versions of itself, as well as act in ways that we could not understand. Such a possibility is far off current capabilities however.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

As the owner of a vast amount of public data the UK Government stands to potentially be a significant catalyst for the development of AI in the UK. The Government could see considerable benefits from investing in AI systems that can cross reference data held by multiple Government departments, for example allowing HMRC and DWP to compare data to identify fraudulent or erroneous benefit applications.

As the application of AI has only just started to become significant, it is perhaps too early to know where regulation may help or hinder the development of this technology. Whilst it is important that personal and sensitive data is kept protected, regulation in this space must be carefully examined so as not to blanket across industries and disrupt the legitimate and responsible application of AI in industries where it will deliver substantial benefits to society. However, it is important that AI be used to benefit humanity, what Fujitsu calls Human Centric, rather than allowed to develop beyond our control. Whilst this AI capability is still somewhat far away, regulation may focus on ensuring that the design and programming of AI does not allow these systems to become autonomous in unintended ways.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

The European Union, in particular the European Parliament, has engaged in a significant amount of work regarding Artificial Intelligence and Robotics. These topics are often analysed together in the EU due to the close relationship between the two (as referenced in the introduction to this response). The papers highlighted below provide some insight into the work that has been done on this topic in the EU.

1. **European Parliament** resolution on a “**Civil Law Rules on Robotics**” (includes AI) – passed by the EP in January 2017
2. **EPRS (STOA)** report on “**Ethical Aspects of Cyber-Physical Systems**” (Intelligent robotics systems)
http://www.europarl.europa.eu/RegData/etudes/STUD/2016/563501/EPRS_STU%282016%29563501_EN.pdf
3. **EPRS** briefing on “**Artificial Intelligence: Potential Benefits and Ethical Considerations**” (broad overview of the issues at hand)

[http://www.europarl.europa.eu/thinktank/en/document.html?reference=IP_OL_BRI\(2016\)571380](http://www.europarl.europa.eu/thinktank/en/document.html?reference=IP_OL_BRI(2016)571380)

4. A German AI institute that was funded through H2020:
<https://ec.europa.eu/digital-single-market/en/innovators/institute-artificial-intelligence-innovation-radar#project-details>
5. **EPRS (STOA):** report on “**Horizon scanning and analysis of techno-scientific trends**”: includes an analysis of how salient various topics are among the wider public: AI is by far the biggest one
[http://www.europarl.europa.eu/RegData/etudes/STUD/2017/603183/EPRS_STU\(2017\)603183_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2017/603183/EPRS_STU(2017)603183_EN.pdf)

6 September 2017

Future Advocacy – Written evidence (AIC0121)

Submission to House of Lords Select Committee on Artificial Intelligence

Cath Elliston, Matthew Fenech, Olly Buston, writing on behalf of Future Advocacy, 6th September 2017

Introduction

1. Future Advocacy is a think tank focused on ensuring that the United Kingdom is best positioned to capitalise on the opportunities and mitigate the risks presented by artificial intelligence (AI). We advocate for smart, forward-thinking policies to responsibly cement the UK's position as a global leader in this field.

2. Last year we released a report called 'An Intelligent Future?', which makes 12 policy recommendations to the Government.⁴⁷³ Since then, we've established a global network of partners in industry and academia to begin to action these proposals. We have also recently written a scoping paper for Sir Tim Berners Lee's Web Foundation on the impacts of AI in low and middle income countries. This involved over 40 interviews in 15 countries with those researching and working with AI and informed a white paper on the topic.⁴⁷⁴ We have also contributed to the House of Commons Science and Technology Select Committee's inquiries on 'Robotics and AI' and Algorithms in Decision-Making'; participated in discussions at United Nations level; and briefed Downing Street staff.

3. In this submission, we outline our position on the sections of the consultation where we think we can add the most value. We are happy to be contacted if you have any questions about our response (see our contact details below), and we are happy for our response to be published in full.

Summary of Recommendations for the UK Government (questions 10 and 11)

1. Develop smart strategies in the face of a fast-changing global economy based on mapping of likely impact of automation by sector, region and demographic group in the UK. This could include supporting businesses to retrain employees and expanding retraining schemes to include platform economy workers.
2. Ensure education systems are adapted to not only develop essential STEM and coding skills but also creativity and interpersonal skills which will be less automatable in the longer term; support corresponding initiatives that encourage underrepresented sectors of society (including women and ethnic minorities) to receive training in AI development and deployment.
3. Retain the safeguards outlined in the EU General Data Protection Regulation to prevent important decisions being taken by algorithms

⁴⁷³ Available at: futureadvocacy.com/s/An-intelligent-future-3.pdf

⁴⁷⁴ Available at: http://webfoundation.org/docs/2017/07/AI_Report_WF.pdf

without human oversight (the “the right not to be subject to a decision when it is based on automated processing”) in the upcoming UK Data Protection Bill. This should be part of a broader ‘new deal on data’ between citizens, businesses and government with policies around consent, privacy, accountability and transparency to give people more control.

4. Ensure that the migration policy in place following Britain’s exit from the European Union will still allow UK-based companies and universities to attract the brightest and best AI talent from all over the world.
5. Scale up and widen the reach of initiatives such the British Growth Fund and the London Co-Investment Fund to support startups, scale-ups and SMEs, relieving the pressure to be immediately commercially successful and allowing increased innovation and commercial risk-taking.
6. Implement regulatory frameworks or legislation that ensure that AI-based systems clearly disclose that they are not human, to increase public trust in the use of AI.
7. Undertake a public information campaign outlining a) the current and imminent uses of AI, b) limitations and risks of its use such as in the context of social media filter bubbles, and c) what the public can do to mitigate these risks.

Artificial Intelligence and Growth (questions 1 and 10)

4. Defining Artificial Intelligence (AI) is difficult, not least because ‘intelligence’ itself is so difficult to define. In this submission we will use an inclusive definition of intelligence as ‘problem-solving’ and consider ‘an intelligent system’ to be one which takes the best possible action in a given situation.⁴⁷⁵

5. AI is already enabling a wave of innovation across many sectors of the global economy. It helps businesses use resources more efficiently (e.g. through automated planning, scheduling, and optimized workflows, supply chains, and logistical pathways) and even enables entirely new business models to be developed, often built around AI’s powerful ability to interrogate large data sets. While McKinsey’s June 2017 discussion paper reported that ‘AI adoption outside of the tech sector is at an early, often experimental stage’, it is being deployed increasingly across industries ranging from manufacturing and ecommerce to the financial services.⁴⁷⁶ We look forward to reading the recommendations of review by Dame Wendy Hall and Jérôme Pesenti on growing the AI sector further.

6. There are two key barriers to a flourishing AI sector in the UK that this review should take into account. In our conversations with CEOs and CTOs about their vision for AI development in the UK they frequently raised concerns that the UK appears to have less of an ‘appetite for commercial risk’ when it comes to

⁴⁷⁵ Russell, S. J., and Norvig, P., (1995) *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ: Prentice Hall.

⁴⁷⁶ Allas, T., Bughin, J., Chui, M., Dahlström, P, Hazan, Henke, N., E., Ramaswamy, Trench, M. (2017). ‘Artificial Intelligence: The Next Digital Frontier?’ McKinsey Global Institute (retrieved from <https://mckinsey.com>, accessed 22 August, 2017)

investment decisions, compared with countries like the United States, for example. It is the experience of many startup founders that potential investors are looking for returns on investment that may be very difficult to guarantee or confidently predict with experimental technology such as AI and robotics. **The UK Government should dedicate funds to supporting startups, scale-ups and SMEs, such that the pressure to be immediately commercially successful is relieved, and allowing increased innovation and commercial risk-taking. In this regard, funding models such as the British Growth Fund and the London Co-Investment Fund have been very helpful to existing start-ups, and should be scaled up and replicated throughout the country.**

7. Furthermore, in all our conversations, concerns about the effects of Brexit on the ability to recruit and retain researchers and other experts in AI were mentioned. In the current climate of uncertainty, there has already been a sharp decline in EU applications to UK tech jobs.⁴⁷⁷ There are 180,000 EU workers in the tech sector but the UK government is yet to confirm new visa rules for EU workers after Brexit. If these workers left the UK it would tear open the already vast 'skills gap'. **During Brexit negotiations and following Britain's exit from the European Union, the government should ensure that a migration policy is in place that will still allow UK-based companies and universities to attract the brightest and best AI talent from all over the world.**

Automation and Inequality (questions 4 and 6)

8. While AI can speed up commercial processes significantly, these savings may come at a price for some employees. There have been numerous predictions about the effects of automation on the job market; PwC predicted in March that up to 30% of UK jobs will be at high risk of automation by the early 2030s, and Deloitte and Oxford University put the figure at 35% in the next 20 years in 2014. Sectors such as 'Manufacturing', 'Wholesale & Retail' and 'Transportation & Storage' are consistently judged to be at highest risk from automation. 'Health & Social Work' is judged to be less automatable.^{478,479} These calculations are informed by a shift in focus to thinking about jobs as a collection of tasks, and therefore focusing on the automatability of individual tasks, rather than whole

⁴⁷⁷ Ram, A. 'Sharp drop in EU job applicants to UK tech industry', The Financial Times, available at <https://www.ft.com/content/8360ed4a-7116-11e7-aca6-c6bd07df1a3c> (accessed 25th August, 2017)

⁴⁷⁸ Berriman, R. and Hawksworth, J. (2017) 'UK Economic Outlook' (retrieved from <https://www.pwc.co.uk/economic-services/ukeo/pwcukeo-section-4-automation-march-2017-v2.pdf> accessed 22 August 2017)

⁴⁷⁹ Frey, C. and Osborne, M. (2014) 'Agiletown – The relentless march of technology and London's response' (retrieved from <https://www2.deloitte.com/uk/en/pages/press-releases/articles/automation-and-industries-analysis.html> accessed 22 August, 2017)

jobs. This is now the accepted standard in this area of research.⁴⁸⁰ Professions which involve processing large amounts of data or routine tasks are classified as vulnerable to automation. This puts junior roles in high-skill professions at risk, like junior lawyers for example.⁴⁸¹ Roles that require creativity, lateral thinking, interpersonal skills, caring, and adaptability are less likely to be at risk.⁴⁸²

9. Of course, any automation estimates must be put in the context of job creation. It is likely that further development and wider implementation of AI will create whole new categories of jobs that we cannot currently envisage. Social media managers and app developers did not exist as employment options at the turn of the century.⁴⁸³ In Canada, AI and machine learning job opportunities, as a share of all job opportunities, have grown by nearly 500 percent since June 2015.⁴⁸⁴ However, there is a risk that the number of jobs that are automated will still vastly outweigh those which are created. The Economist makes a stark comparison: 'at its current pace, by July 2018 retailing will have shed three times as many jobs as Amazon is due to create'.⁴⁸⁵

10. Another important challenge is that job creation will likely be concentrated in high-skill professions with few benefits for low-skilled and medium-skilled workers.²⁵ Reskilling safety nets are vital. At Accenture, 17,000 jobs were automated but no-one lost their job, a feat that CEO of financial services Richard Lumb attributed to reskilling.⁴⁸⁶ But the retraining opportunities for lorry drivers for whom the prospect of automation looks to be accelerating are less clear. In August the government announced that small convoys of partially self-driving lorries will be trialled on major British roads by the end of next year which poses

⁴⁸⁰ Arntz, M., T. Gregory and U. Zierahn (2016) 'The risk of automation for jobs in OECD countries: a comparative analysis', OECD Social, Employment and Migration Working Papers No 189 (available at http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en?crawler=true)

⁴⁸¹ Croft, J. (2017) 'Artificial intelligence closes in on the work of junior lawyers', Financial Times, available at <https://www.ft.com/content/f809870c-26a1-11e7-8691-d5f7e0cd0a16> (accessed 23 August, 2017)

⁴⁸² 'An Intelligent Future? Maximising the opportunities and minimising the risks of artificial intelligence in the UK', 25 October, 2016. Available at <https://www.futureadvocacy.com/s/An-intelligent-future-3.pdf>

⁴⁸³ Bowers, E., (2017) '10 Jobs Created by Tech That Didn't Exist 10 Years Ago', The Nasstarian, available at <http://blog.nasstar.com/10-jobs-created-by-tech-that-didnt-exist-10-years-ago/> (accessed on 5 April, 2017)

⁴⁸⁴ Zubairi, A. (2017), "Report: Canadian job opportunities in AI have grown by nearly 500%", Betakit, available at <http://betakit.com/report-canadian-job-opportunities-in-ai-have-grown-by-nearly-500/> (accessed on 30th August, 2017)

⁴⁸⁵ The Economist (2017) 'The decline of established American retailing threatens jobs' available at <https://www.economist.com/news/briefing/21721900-love-affair-shopping-has-gone-online-decline-established-american-retailing> (accessed 4 September 2017)

⁴⁸⁶ Brinded, L. 'Automation killed 17,000 roles at a huge tech and services firm — but no one actually lost their job' Business Insider, available at <http://www.businessinsider.com/accentures-richard-lumb-davos-interview-robots-jobs-skills-leadership-training-2017-1?r=UK&IR=T> (accessed 25 August, 2017)

a risk to the haulage and logistics industry's 2.2 million employees.^{487,488} The impact of this may be lessened given the shortage of HGV drivers, which the Road Haulage Association estimates at 60,000 drivers, but the disruption in this sector among others is certainly worth mapping and planning for.

11. It is heartening to see that this year's Queen's Speech promised to 'ensure people have the skills they need for the high-skilled, high-wage jobs of the future, including through a major reform of technical education'. This should encompass a drive on STEM skills and coding in schools, but must also encourage creativity, adaptability, caring and interpersonal skills which will provide a crucial comparative advantage for humans over machines over a longer timeframe.⁴⁸⁹ Particular focus must be on jobs where retraining opportunities may be limited, and equality of access to retraining programmes regardless of gender, ethnicity, and socioeconomic background must be ensured. The social and psychological impact of being out of work for people of all ages and backgrounds should not be underestimated. The Conservative Party's manifesto commitments to introduce a new right to request leave for training for all employees, and to establish a state-funded 'national retraining scheme' in tandem with this new right, are welcome and we look forward to more detail being published. The growing number of platform economy workers should also have access to this retraining scheme. Uber now operates in over 20 UK cities, employing 25,000 people in London alone.⁴⁹⁰ The company's stated goal is to replace their drivers entirely, which would drive down costs and accident numbers, but also jobs.⁴⁹¹

12. Another intervention being mooted (most prominently by Bill Gates) to mitigate the potential effects of automation is the introduction of a so-called 'robot tax'. This may constitute a source of funding to support employee retraining programmes, as suggested by the European Parliament earlier this year.⁴⁹² Furthermore, 'robot taxes' might provide a solution to the potential problem that reduced employment will lead to reduced income tax and National Insurance revenues; these two tax types are, along with VAT, the largest sources

⁴⁸⁷ Campbell, B. 'Self-driving lorries to be tested in UK next year' The Financial Times, available at <http://www.bbc.co.uk/news/technology-41038220> (accessed on 29th August, 2017)

⁴⁸⁸ The Road Haulage Association, "About the Road Haulage Industry", available from <https://www.rha.uk.net/policy-campaigning/haulage-industry> (accessed on 30th August, 2017)

⁴⁸⁹ Autor, D. H., (2015). Why are there still so many jobs? The history and future of workplace automation. *The Journal of Economic Perspectives*, 29(3), 3-30.

⁴⁹⁰ Titcomb, J. (2016, 2 June) Majority of Uber drivers in London work part time, study says. *The Telegraph* (retrieved from <https://telegraph.co.uk>, accessed 6 October, 2016).

⁴⁹¹ Newman, J. (2014, 28 May) Uber CEO Would Replace Drivers With Self-Driving Cars. *Time* (retrieved from <https://time.com>, accessed on 6 October, 2016).

⁴⁹² European Parliament (2017), "Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))", available at <http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&mode=XML&reference=A8-2017-0005&language=EN> (accessed on 29 March, 2017)

of revenue for the UK Government, together accounting for almost 60% of total tax revenue according to the Institute for Fiscal Studies.⁴⁹³ There is more work to be done on the practicalities of a “robot tax” that also fosters innovation. But certainly there is a strong logic to the idea that income taxes unfairly disadvantage human labour and act as an unnecessary further incentive to automation. Universal Basic Income is another proposed solution to AI-related job displacement that is gaining traction. Ultimately, we support a taxation model that results in fairer redistribution of the wealth that these technologies will create, rather than having this wealth concentrated in the hands of a few commercial entities who own robots and other automated technologies.

Data: Bias, Transparency, Privacy (questions 4, 7, 8 and 9)

13. AI systems can be trained on biased data which throws up a series of ethical questions. Algorithms may reflect the characteristics of data, including any biases, in the actions they recommend and models they create. This is especially concerning as these systems look set to play an increasing part in high-stakes decision-making, in domains such as hiring, parole, and insurance. Biases can occur when the data available is not an accurate reflection of what it is taken to represent, which could be a result of inaccurate measurement methodologies, incomplete data gathering or other data collection flaws. This type of bias can sometimes be prevented by ‘cleaning the data’⁴⁹⁴ or making the data collection process more robust.

14. Bias may also occur when a process being modelled itself exhibits unfairness. For example, men may be prioritised in job applications if the data used to select candidates was gathered from an industry that systematically hired men over women. A biased algorithm meant that Google ads promising help getting jobs paying more than \$200,000 were shown to significantly fewer women than men.⁴⁹⁵ Addressing this kind of bias may require a combination of common sense along with more complex and political kinds of interventions to establish an ethical framework and avoid reinforcing unfair stereotypes and inequalities. As the White House ‘AI Now’ paper outlined, while communities are doing great work on these issues, there is not yet a consensus on how to detect biases.⁴⁹⁶

15. Concerns about bias are compounded by the severe lack of diversity in the AI field which raises the prospect that bias may be considered less of a problem or

⁴⁹³ Pope, T., and Waters, T. (2016) ‘A Survey of the UK Tax System’, Institute for Fiscal Studies, available at <https://www.ifs.org.uk/bns/bn09.pdf>

⁴⁹⁴ Data cleaning refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of a data set and then replacing, modifying, or deleting them.

⁴⁹⁵ Spice, B., (2015) “Questioning the Fairness of Targeting Ads Online.” Carnegie Mellon University, available at <http://cmu.edu> (accessed 10 March, 2017)

⁴⁹⁶ AI Now (2016), “The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term: A summary of the AI Now public symposium”, The White House and New York University’s Information Law Institute, July 7th, 2016, available from <https://www.artificialintelligencenow.com> (accessed 2nd March, 2017)

may not be identified when it occurs. Kate Crawford has written compellingly about what she terms ‘artificial intelligence’s white guy problem’, whereby a lack of representation limits the perspectives and experiences of AI’s creators, leading to a greater possibility of “like me” bias. In our work with the Web Foundation, we interviewed many developers and AI experts from diverse backgrounds and they were unanimous in their opinion that the AI community needs to be more inclusive. Women are still dramatically underrepresented in coding, STEM subjects and related careers. A-Level data this year showed that only 9.8% of those who completed a computing course were girls.⁴⁹⁷ Despite a number of commendable initiatives, sustained effort will be needed to reverse the low take up to combat bias amongst other things. **The UK Government should initiate and/or support initiatives that specifically encourage underrepresented sectors of society (including women and ethnic minorities) to receive training in AI development, deployment, application and interpretation.** Such initiatives could take the form of subsidised education, targeted information campaigns and/or mentorship schemes, for example.

16. It can be even more difficult to address and resolve bias when the process by which an algorithm arrived at its output is not discernable. This is commonly referred to as a ‘black box’ issue. The effects of this problem range in severity. In healthcare, for example, an erroneous treatment recommendation as a result of an algorithm failing to take a patient’s ethnicity into account (perhaps because the training dataset was enriched with data from one particular ethnic group only) could lead to serious harm. The creators and deployers of AI algorithms should be able to provide an explanation that people can understand as to why decisions have been made. This reduces the likelihood of bias and unjust or incorrect decision making. It also empowers those at the receiving end of these decisions to challenge potentially erroneous outcomes. Even when the decision-making is relatively low stakes, ‘black-boxing’ sets a dangerous precedent. In view of the immense social importance of algorithmic transparency, private and academic developers of AI tools should be supported in their research into opening up the black box. **The EU General Data Protection Regulation (GDPR) provides safeguards for individuals against the risk that a potentially damaging decision is taken without human intervention. These safeguards must absolutely be retained in the upcoming UK Data Protection Bill.**

17. The benefits arising from AI’s ability to interrogate big data may be enormous, but there is also the risk that information people would rather have kept confidential will be revealed. Certain forms of data, such as commercial and medical information, are collected and stored under conditions of anonymity. However, advances in AI make anonymity increasingly fragile and it may become

⁴⁹⁷ Wakefield, J. (2017) ‘Very Few Girls took computing A-Level’ available at <http://www.bbc.co.uk/news/technology-40960427> (accessed 24th August 2017)

increasingly possible to re-assign identity to particular sets of information because of AI's ability to cross-reference between vast quantities of data in multiple data sets.⁴⁹⁸ These developments worsen existing concerns about privacy and raise new ones.

18. AI could also be used to undermine the notion of consent. Given the swift advances in data analytics it is impossible to imagine all the uses data may serve in the future, making it hard to assure the protection of data subjects. Google Deepmind's partnership with the Royal Free Hospital to better diagnose kidney disease with machine learning necessitated the transfer of 1.6 million patient records. The Information Commissioner (ICO) raised concerns about how much the patients knew about this use of their data earlier this year.⁴⁹⁹

19. Advances in AI could worsen existing tensions over how data is used by the government and public bodies. When it was being debated, the Digital Economy Bill was criticised for its 'thin safeguards' regarding the sharing of publicly held data, and for a lack of precision in defining data sharing, though these were addressed by the House of Lords.⁵⁰⁰ Over time these issues could become more important, as the amount of information held about us grows and the ability to analyse it improves.

20. We need a 'New Deal on Data' between citizens, business, and governments.⁵⁰¹ This is in the interests of business and government as it will build trust. If we do not have a deeper public debate we risk undermining public confidence in this new technology, sparking opposition to its uptake. The government needs to ensure all stakeholders can raise concerns in an open and constructive manner. Greater clarity is needed about who collects what, and for what purpose. People need to understand the rights of various parties and how to access information about how their own personal data is stored and used. Public debate should also focus on the uncertainties around how data might be used in the future. We organised a roundtable discussion with Amnesty International in September, where we built on our proposal for New Deal on Data in conjunction with key stakeholders, such as the Open Data Institute, Privacy International, Nesta and the Royal Society.

⁴⁹⁸ See Ohm, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA law review*, 57, 1701

⁴⁹⁹ Hern, A. (2017) 'Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind' *The Guardian*, available at <https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act> (accessed 25th August, 2017)

⁵⁰⁰ Fiveash, K. (2017), "Digital Economy Bill passes through House of Lords and will soon be law", *Arstechnica UK*, available at <https://arstechnica.co.uk/tech-policy/2017/04/digital-economy-bill-passes-through-house-of-lords-will-soon-be-law/> (accessed on 30th August, 2017)

⁵⁰¹ The case for a 'new deal on data' has been made in the USA by Alex "Sandy" Pentland. See for example *Harvard Business Review* (retrieved from <https://hbr.org/2014/11/with-big-data-comes-big-responsibility> accessed 10 October 2016).

21. As part of the New Deal on Data, we support the call of the British Academy and the Royal Society to establish both a set of high level principles around data governance, and a high-level body to 'steward the evolution of the governance landscape as a whole', and conduct expert investigation into issues posed by these new technologies and their future consequences.⁵⁰²

Public Perception and Influence (questions 5 and 8)

22. Future Advocacy are conducting a series of YouGov polls to determine UK public attitudes to AI. In our 2016 poll we asked whether the UK considers AI to be 'more of an opportunity or a risk'. 30% of men and 29% of women considered it 'more of a risk' and a further 26% of women and 16% of men said that they didn't know. This suggests a level of both suspicion and uncertainty in the public's current perception of the technology, influenced perhaps by cultural narratives in television, film and the press where AI tends to be portrayed in a dystopian light. At the same time, 49% of respondents claimed that they were not worried about AI taking their job, compared to just 8% who were either 'fairly worried' or 'very worried'. It follows that while AI might be distrusted in the abstract, the public tend to dismiss the prospect of their own jobs being automated. This is at odds with predictions, quoted at the start of this submission, which show that around 30% of jobs are at high risk of automation. Our poll this September will show if public attitudes have changed in the last year.

23. There are particular areas where the public's understanding of AI can have detrimental effects. AI processes online often operate in hidden ways that exempts them from public scrutiny. For example, automated accounts known as bots which operate on social media are sometimes indistinguishable from humans. These accounts were used to express vocal political support on social media in the run up to the UK election. In fact, they generated one in eight tweets about British politics.⁵⁰³ While the precise impact of this activity on the result is hard to measure, if an automated tweet is not recognised as such, it certainly undermines the principles of informed decision-making that underpin the democratic process.

24. Furthermore, the information that people see on websites and social media is to a great extent determined by AI algorithms that the public may not be aware of. AI-powered voter profiling allows the personality and opinions of voters to be analysed through data analytics and machine learning to effectively target certain demographics with selective information. Algorithms are also used to tailor the information users see online based on their likes and dislikes. This can reinforce opinions that an individual already holds, by reducing the amount of information they see that challenges their viewpoints. **The public should be**

⁵⁰² British Academy and Royal Society (2017) Data management and use: Governance in the 21st century

⁵⁰³<http://www.ox.ac.uk/news/2017-05-31-labour-dominating-twitter-conversation-uk-election-campaign-says-study>

made aware of how AI is being used online so that they are less likely to have their opinions artificially manipulated. Furthermore, there should be regulations which state that AI systems must clearly disclose that they are not human, as Oren Etzioni, CEO of the Allen Institute for AI has recently suggested.⁵⁰⁴ This may help to circumvent the problem of bots influencing election results and improve general public understanding of AI.

25. Another potentially damaging perception of AI is the notion that it is infallible. It is important to recognise the limitations of data analysis. Correlation does not equal causation. AI is capable of recognising patterns, and large diverse datasets can throw up many patterns indeed. Some are meaningful, others are not. This should be borne in mind as our use of these insights increases, especially when used to inform public policy. Google's ability to predict flu outbreaks, for instance, failed after what seemed like strong initial successes.⁵⁰⁵ If even experts are tempted to over-rely on AI-based decision-making, an uninformed public may feel even less empowered to challenge algorithmic decisions made about them.

26. AI is already affecting the public and their decisions, albeit in ways that are often invisible and difficult to measure. At the very least we need to inject practicalities and realism into the discourse around AI which tends to be speculative and far removed from the everyday realities of public life. The government has a crucial role to play, particularly around the election cycle to uphold democracy. Voter registration drives could be accompanied with practical information on the ways in which AI can be used to sway voters. Urging sites like Facebook - who placed newspaper adverts on how to spot fake news - to do more around these issues is also vital.

27. Future Advocacy will be conducting research in order to:

- Gauge the attitudes of the public in relation to particular high-stakes issues which could inhibit the roll out of AI or negatively impact swathes of the population
- Better understand the risks of job displacement (consequent to increased automation of roles and tasks) within specific sectors of industry and in different constituencies.

6 September 2017

⁵⁰⁴ Etzioni, O. (2017) 'How to Regulate Artificial Intelligence' The New York Times available at <https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html?mcubz=0> (accessed 5th September, 2017)

⁵⁰⁵ Lazer, D. and Kennedy, R. (2015) 'What We Can Learn From the Epic Failure of Google Flu Trends' Wired, available at <https://wired.com> (accessed 11 October, 2016).

Future Intelligence – Written evidence (AIC0216)

Introduction:

1. This submission is made on behalf of Future Intelligence, the scientific research arm of Cyber Security Research Limited. CSR's directors include some of the UK's leading experts on cyber security⁵⁰⁶. CSR's Chair is Peter Warren, an award-winning technology journalist. Warren has co-authored two books⁵⁰⁷ on cyber security and an influential report on AI, the Internet of Things (IoT) and Big Data that was presented to the EU Digital Agenda team and the French Senate⁵⁰⁸. Cooley and CSR with the Institute of Engineering and Technology held an international AI conference in London in May 2017⁵⁰⁹. Peter Warren presents a monthly hour-long radio documentary on the ramifications of technology, PassWOrd, on Resonance FM, the alternative London radio station www.resonancefm.com. He is the editor of the technology news websites www.futureintelligence.co.uk and www.csri.info

2. The authors respectfully ask the noble Lords to recognise that Artificial Intelligence cannot be considered in isolation from other associated technologies, all of which are dependent on each other for their development. Thus, AI is dependent on the collection and effective use of big data, and autonomous vehicles, including battlefield drones and robots, are all dependent for their effective functioning on AI systems.

What is the current state of artificial intelligence and what factors have contributed to this?

3. AI is pervasive in banking, insurance and finance, and in automating and targeting customer service and marketing. Algorithms can calculate, generate leads, assess claims, bid, buy and sell far faster than the human mind and deliver better results with no labour costs. AI is in farming, stock markets and manufacturing. Already corporations exist that make money for their human owners without any human intervention. (e.g. DACs Decentralised Autonomous Corporations,) According to Professor Joanna Bryson of the Universities of Bath and Princeton, AIs have recently been trained to lip-read and there are few fields where it cannot be applied.⁵¹⁰

⁵⁰⁶ <http://www.csri.info/about-csri/>

⁵⁰⁷ *Cyber Alert*, Streeter and Warren, Vision, 2005, *Cyberwarfare – all that matters*, Streeter and Warren, Hodder and Stoughton, 2014

⁵⁰⁸ *Can we make the digital world ethical?* Streeter, Warren and Whyatt, Netopia, 2014

⁵⁰⁹ <http://www.futureintelligence.co.uk/events/>

⁵¹⁰ Do we need a kill switch? with Mark Deem and Professor Joanna Bryson at Future Intelligence AI conference 2017 <http://www.futureintelligence.co.uk/events/>

How is it likely to develop over the next 5 years?

4. It will extend into logistics, transport, legal services, health and education. Personal use of AI will increase, building on the Siri and Alexa devices embedded in mobile phones and household AI systems. These will become our individual digital agents or avatars⁵¹¹, searching the world-wide web to find what they have assessed we need and bringing those things to our online shopping basket or library. The first widespread commercial use of an avatar was in 1996 with the unsuccessful introduction of the Microsoft 'Paper Clip' in the Windows 97 operating system, that lesson has led to the development of taking avatars such as Siri and Alexa 2012 the Antique Trader's Tophatter website has used avatars⁵¹² -the logical next step is a personal avatar that operates across all websites.

10 years?

5. AI will be used more widely in personal life (see above). We shall see the widespread use of autonomous vehicles and experts at Oxford University predict computerisation will replace humans in around one third of all UK jobs.⁵¹³ AI medical examination systems will work alongside nursing staff,

20 years?

6. "AI" will no longer be a valid term, since the digital and physical worlds will be so deeply interconnected in a global neural network that the notion of separating out a single algorithm will become quaint and irrelevant, according to thinkers such as Professor Nick Bostrom⁵¹⁴ at Oxford University and Ray Kurzweil⁵¹⁵ at Google. So, for example the AI in your personal digital assistant will be part of the mesh of all the other algorithms.

What factors will accelerate or hinder this development?

Technical factors

7. Cyber security: "This is the biggest roadblock to AI adoption right now" – Prof Joanna Bryson, *ibid*. The integrity of the data is essential for the functioning of AI. It is essential that the current crisis in cyber security must be resolved. It should be noted that even in the Critical National Infrastructure the sensor systems have no cyber security and their computer chips are too small to hold the programming requirement for cyber security protection. The Critical National

⁵¹¹ Can we make the digital world ethical? p7 <http://www.netopia.eu/wp-content/uploads/2014/02/Report-Can-we-make-the-Digital-World-ethical.pdf>

⁵¹² Antique dealing avatars <http://www.antiquetrader.com/antiques/new-online-business-model-allows-live-auction-thrill-in-a-virtual-house/>

⁵¹³ Predicted employment trend

http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

⁵¹⁴ <https://www.thalia.de/shop/home/rubrikartikel/ID43939392.html?ProvID=11000522>

⁵¹⁵ <http://singularity.com>

Infrastructure (CNI) includes telecoms, the National Grid, water supplies and air traffic control. This means that AI control of utilities and power stations would not be advisable at the current time. Bryson (ibid.) says that AI can do just about anything. So, the only blocks are the need to establish human primacy over smart machines, overcoming public reservations and determining the legal status and liability of autonomous vehicles and systems.

Societal factors

8. People and policy makers must address a number of legitimate concerns about AI:

- Autonomous weapons can kill without human agency or intervention⁵¹⁶
- Ownership of data, the 'new oil' – given the value of data to the big data pool, people will increasingly demand a return for the use of their data. This will lead to disputes between individuals and companies over the value of crude data. The companies will claim that the value of information is arrived at through its refinement, and that crude data on its own has little value until amalgamated with other data from other sources in the big data pool.
- Responsibility sits at the heart of all of this. If the decisions are left to machines then people abrogate their responsibility. This raises contradictions: a robot doctor may be better as a decision maker than a real doctor – so who in that position should make the decision? We note that military drone operators and those in charge of missile systems are loath to over-ride automated systems. They fear making a mistake for which they will be blamed. (Professor Kevin Warwick 'March of the Machines').⁵¹⁷
- The effect on employment of a widespread move to AI-controlled automation (see below)
- Concerns about privacy, state surveillance, unwanted marketing (junk emails)
- Dangers from hacking by criminals, political activists or state actors to compromise vital systems such as the National Health Service⁵¹⁸ and parts of the Critical National Infrastructure⁵¹⁹.
- Questions of legal liability remain unresolved: this could put a brake on the roll-out of autonomous vehicles. Cooley's Mark Deem explained at our 2017 AI conference (ibid) that this is a complex issue.

⁵¹⁶ <http://www.futureintelligence.co.uk/2017/08/scientists-urge-un-to-ban-robot-weapons/>

⁵¹⁷ March of the machines by Professor Kevin Warwick
<http://www.press.uillinois.edu/books/catalog/67cmc6ff9780252072239.html>

⁵¹⁸ <http://www.futureintelligence.co.uk/2017/05/did-nhs-spending-cuts-open-the-door-to-ransom-attacks/>

⁵¹⁹ <http://www.futureintelligence.co.uk/2013/06/watchdog-warns-on-bts-chinese-whispers/>

- Lack of accountability: how can politicians be held to account for policy decisions that were derived from algorithms? Who or what is to get the praise or blame?
- Lack of ethics and regulation: Google has installed an Ethics Board, but there is no over-arching regulatory body for AI and no consensus about the rules and norms that should be applied. Cooley's Mark Deem believes the law needs to catch up with technological advances. One way to do this is to give AI a legal status as an entity, so that it can become subject to the law of the land. MEPs have tried to enshrine this principle in the EU Civil Law on Robotics⁵²⁰. The IEEE has set up a global ethics team.⁵²¹

These piecemeal approaches could be enhanced if the technology industry could be persuaded to create its own version of the Hippocratic Oath, ideally fused with the Duty of Confidentiality, Duty of Care and Duty of Candour that apply to the medical, spiritual and legal professions. One way this could be introduced in the UK might be through the Worshipful Company of Information Technologists, and through industry bodies such as the CREST, IET and IEEE. The Hippocratic Oath for software engineers was first mooted by Philip A. Laplante of Penn State University in 2004⁵²². He also backs a system of software testing and licensing recommended by Peter Warren, one of this report's authors. Warren argues that software should be tested – like a new drug or food product – and approved before it is released, in a rigorous system like that of the UK National Institute for Clinical Health Excellence (NICE) or US Federal Food and Drug Administration (FDA)⁵²³. Certification of qualified information technologists and software engineers occurs through automated self-testing by the companies who are doing the hiring and firing⁵²⁴. If knowing and understanding the code of conduct and oath formed a compulsory part of the assessment process for certification – with scenarios or problems to test that the trainee understands how it works in practice – that would be a means to embed ethical codes into the training and education of software engineers and would work as an effective brake to the development of unethical software as a benchmark would exist. There is a stark contrast in sentencing policy for computer crime between the UK and the USA. Given the interdependency of crucial systems, interference in one area could have fatal consequences. Penalties should apply both to those creating the weakness (poor or unethical coders and manufacturers) and those exploiting it (hackers and criminals).

⁵²⁰http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU%282016%29571379_EN.pdf

⁵²¹ https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

⁵²² <http://queue.acm.org/detail.cfm?id=1016991>

⁵²³ Recommendations on page 28 of Warren's digital ethics report

<http://www.netopia.eu/wp-content/uploads/2014/02/Report-Can-we-make-the-Digital-World-ethical.pdf>

⁵²⁴ <http://www.tomsitpro.com/articles/programming-certifications,2-274.html>

Democracy

9. The effect on democracy is already being felt: AI permits the automatic generation of “fake news” and abusive content: “trolling” “cyber-bullying” or “hate speech” Some governments pay computer programmers to target their political opponents in this way through social media feeds. Many thousands more people produce fake news and hate speech for money, earning dollars in advertising fees from Google and Facebook⁵²⁵. This impacts democracy, as shown by the United States 2016 election. US security services all agree that Russia’s Vladimir Putin⁵²⁶ should be investigated for his role in manipulating social media users to swing voters’ choice. A British company, Cambridge Analytics, played a part in this process. And a fugitive in London’s Ecuadorian embassy, Julian Assange, released emails through the Wikileaks whistleblowing platform which also influenced the USA 2016 election. In the future, AI itself should be able to kill off fake news in social networks. However, it will take some time to reach that stage and meanwhile it is possible to mis-train and distort the model with fake news. Microsoft was obliged to disable its chat function when its bot started using racist and sexist language. Facebook has responded to Germany’s draconian new legislation by hiring the Correctiv! journalists’ collective to identify and remove fake news from its social network in order to avoid large fines⁵²⁷.

10. In response, the UK Director of Public Prosecutions has signalled a new tough approach to fake news and online hate crime, and launched a public consultation.⁵²⁸

NATO and the European Union have combined forces to build a counter-propaganda centre in the Finnish capital Helsinki to neutralise hybrid threats from Russia (“hybrid” means actual cyber-security breaches and “psy ops” or psychological operations through social media).⁵²⁹

The European Union’s Fundamental Rights Agency has commissioned its own AI solution: a bot called “Franny” that seeks out abusive words online and responds to them automatically with a reprimand and friendly emojis.⁵³⁰ There is a need to establish trust through libraries of record, institutions of record, sensors of record and journals of record. We recommend a rating system for websites in which their content’s trustworthiness – a truth co-efficient - is given a score from one to ten. This can be indicated on every site like the tiny green padlock that appears as a security certificate, or the yellow canary symbol that is used to

⁵²⁵ <https://www.wired.com/2017/02/veles-macedonia-fake-news/>

⁵²⁶ <http://www.futureintelligence.co.uk/2017/04/fake-news-threat-to-european-election-warning/>

⁵²⁷ <http://www.telegraph.co.uk/technology/2017/06/30/germany-fine-facebook-youtube-50m-fail-delete-hate-speech/>

⁵²⁸ http://www.cps.gov.uk/news/latest_news/cps_publishes_new_social_media_guidance_and_launches_hate_crime_consultation/

⁵²⁹ http://www.nato.int/cps/en/natohq/news_143143.htm

⁵³⁰ <http://fra.europa.eu/en/news/2016/frf-hackathon-and-winner-franny-fundamental-rights-bot>

indicate that the website owners will never share users' details with security services or law enforcement. It can be seen, then, that the potential for damaging the democratic process is being addressed in various ways already by the UK and her allies, and by the technology industry and communities themselves. These different approaches require co-ordination, rigorous evaluation and pooling of resources.

The most challenging point relating to AI and democracy is the lack of choice that is offered to the population at large about the adoption of technology. It is to say the least undemocratic. It is noteworthy that those being asked to embrace the smart-city are not asked if they want it. That those whose jobs are being replaced are not being canvassed about that. While the Brexit decision was in some part caused by resentment about competition from lower paid workers, similar democratic rights over the adoption of robots and AI systems that remove jobs are not proffered. Nor, is a vote offered on whether controls should be introduced on potential enhancements to individuals based upon income that would give them massive advantages over their peers.

Jobs

11. In the UK there is a severe skill shortage in the relevant disciplines. For example, Engineering UK estimates that a further 1.8 million engineering graduates are needed. Ben Rossi quotes this statistic in *Information Age*, and notes: "In 2016, only 5,600 students studied computer science at A level (only 600 of them female) versus 31,000 studying sociology."⁵³¹

Research funding which until now has been provided through the European Union, such as Horizon 2020, will no longer be available to UK institutions after Brexit in March 2019⁵³².

Employers (companies and local and national government agencies in the public sector) who switch manufacture of products or delivery of services from all-human to total automation will probably provoke industrial strife and possibly civil unrest. However, if AI is introduced gradually - as a means of making human workers' jobs easier or less repetitive - and mass lay-offs are avoided, this seems unlikely. Many of today's workers use an AI such as Siri in their smartphones or Nest in their homes. They find them helpful and may anthropomorphise them as "friends"⁵³³. This attitude makes them less likely to turn Luddite.

As this is likely to happen we need to have device sanctity, and that needs to be guaranteed and protected. The state cannot be allowed to have any control over personal devices (such as mobile phones or tablets) or the state becomes pervasive, intrusive and authoritarian. If states achieve power over the loyalty of the devices that we pay for then the world of George Orwell's 1984 is only a

⁵³¹ <http://www.information-age.com/britain-can-solve-critical-digital-skills-crisis-123467388/>

⁵³² <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>

⁵³³ <https://academic.oup.com/jcr/article-abstract/44/2/414/2939535/Products-as-Pals-Engaging-with-Anthropomorphic?redirectedFrom=fulltext>

heartbeat away. The devices will not be trusted and the future model of AI and the internet will crumble.

Is the current level of excitement which surrounds AI warranted?

12. There is a tendency to hype AI at the moment, so that it would appear that everything has AI in it. In fact, it is simply a recognition of how many systems already had AI in them. Companies have been using algorithms for a long time now and are suddenly being alerted to the fact that they have “an AI component” in their product and are thus badging their system as ‘AI inside’ just as Intel did in the mid-1990s. That said AI is increasingly being seen as the solution to handling the big data problems and complexities that have emerged as the result of the internet explosion and the attendant issues such as computer security. Paradoxically, AI would not have been possible also without the amount of big data being generated by modern communications technology.

Yet it does signal a new industrial revolution and it comes⁵³⁴ with a cultural revolution, since human activity will become confined to creative and caring tasks and the machines will provide the infrastructure of work and life.

How can the general public best be prepared for more widespread use of artificial intelligence?

13. AI should be made more transparent, for example by clearly labelling its presence wherever it occurs and allowing the user to opt out, in the same way that cookies are labelled now.

14. Privacy concerns can be addressed by a system, suggested by Cooley’s Mark Deem, of standard terms and conditions (Ts and Cs). Different versions could be labelled Bronze, Silver and Gold and used to kitemark online services, apps, etc and where individual companies may also add its own riders or exceptions. The companies should underline where they are departing from an accepted norm and should prominently flag that up so that people can see what they are doing.

15. The skills shortage and tendency to anthropomorphise AI, noted above, mean that a nationwide campaign of public awareness-raising is needed, to ensure that people understand the implications of AIs in their world. Their role should be thoroughly explained, as should the mechanisms that have been put in place to control them. The public should be told what they need to do to educate themselves to ensure that they are always in a position of primacy. Technology should be made democratic. It should be illegal for enhancements to be made that are not available to the mass of the population. This has not been possible in the past and it should not be in the present or future.

Kings in the past could be trained in weaponry, could be given a better diet than their fellows, could be given better tuition, better armour, but those benefits would only last for the lifetime of that individual and did not lead to permanent divisions. Edward II succeeded Edward I and was nothing like his father. AI has

⁵³⁴ The internet chapter in *The future of business*, Whittington, Fast Future Publishing, 2017

the potential to allow one class to artificially separate itself permanently from another. It also offers the possibilities of cloning and through transhumanism⁵³⁵, possibly for eternal life. If AI and enhancement is allowed to run amok, society will be governed by superior beings and change will be imposed not accepted. This should be explained to the public – the development of a super intelligent society should be prevented in the interests of humanity. Otherwise the notion of democracy will become a total farce.

16. Cyber-security risks will grow exponentially as the Internet of Things is rolled out, and more investment is needed in encryption, fraud detection, insurance, training and education - especially for the older generations. It is unhelpful when a Government Minister⁵³⁶ decries encryption as a means for concealing crime or terrorism. "Without encryption Britain cannot do business and all our intellectual property and trade secrets are at risk" says Peter Warren, Chair of the Cyber Security Research Institute. "We should encrypt by default".

17. Large-scale job losses caused by a move to AI require mitigation. MEPs recommended a Universal Basic Income in the debate on the Civil Law on Robotics but this was voted out of the final rules. (see below)

Still, where a community relies heavily on one big employer (such as a government department) that is about to be automated, the government should consider investing in re-training and social opportunities.

18. The well-publicised gender inequality and pay gap⁵³⁷ in the technology industries means that the next decade could see a major shift in the role of women in society. Just as in the 1950s when fighting men returned from World War 2, the 2020s could see a mass return of women (and older people of both sexes) to unpaid homemaking and caring roles, with the few remaining jobs held by highly-skilled and highly-paid young men. For Ms Average there may be few options: low-paid, boring work for herself, a robot nanny for the children and a network of AI sensors watching over the elderly parents in automated sheltered housing. As added benefits, children would grow up with enhanced social skills and senior citizens would grow old with dignity, feeling valued in a family setting. However, in this scenario, seventy years of progress towards female emancipation would be wiped out. The potential contribution of women to industry, commerce and public life would be lost.

19. An alternative would be a government-backed drive to re-skill the existing female workforce and to make Advanced Level Mathematics and computer programming compulsory subjects for all students of both sexes who aim to

⁵³⁵ <http://www.zoltanistan.com/TranshumanistParty.html>

⁵³⁶ <http://www.telegraph.co.uk/news/2017/07/31/dont-want-ban-encryption-inability-see-terrorists-plotting-online/>

⁵³⁷ Gender inequality and pay gap in technology

<http://www.opusrecruitmentsolutions.com/blog/uk-women-in-tech-what-s-the-situation-in-2017--blog-76231013458>

enter any higher education course, as it is in Russia⁵³⁸

20. Before the first industrial revolution, Britain was an agrarian society with cottage industries. What if we should turn the clock back to 1750 instead of 1950? Working from home, buying and selling online, delivering education or health advice or making bespoke goods using AI and 3D printing – these are all options for those who have lost jobs to automation, or never had a proper job. The popularity of the Maker Faires movement⁵³⁹ in parts of the UK shows that this way of life has the potential to become more widely adopted.

4. Who in society is gaining most from the development and use of AI and data?

21. Hedge fund managers, big corporations, financial institutions and the companies that own the algorithms: search engine companies, social media groups, and technology companies like Microsoft, IBM, BT, O2, financial institutions, cyber security companies, utilities, large farming organisations. AI raises some interesting challenges, such as remuneration for instance – increasingly people are being provided with infrastructural technological services. Search engine services and social media platforms are making the users unwittingly work for the companies for free, provide the companies with their personal data, their search patterns etc. The companies are not taxed on this activity, yet it counts towards the estimation of a company's share value. (We will discuss this more in the next section).

Some doctors and hospital managers are using AI to speed up diagnosis and streamline administration.⁵⁴⁰

Who is gaining the least?

21. People who are losing their jobs, or whose jobs are becoming precarious and casualised, because processes are being automated. For example, in July 2017, 910,000 UK workers are on zero-hours contracts.⁵⁴¹ This is already 101,000 more than in 2016. Many automated industrial processes are programmed to deliver only a limited number of products that can be delivered on a "just in time" basis to the point of sale. (e.g. in one case reported to the report's authors, a packer working at the Suffolk sushi factory Ichiban was told to go home at 2am, two hours into what should have been an eight-hour shift. The reason: the batch was finished.)

22. Users of technology systems such as social media are being pushed into

⁵³⁸ In Russia Advanced Mathematics is compulsory for all students

https://www.emis.de/data/projects/reference-levels/EMS_RUSSIA.pdf

⁵³⁹ Make Faires for teaching coding with creativity <http://techmog.com/5686/hacked-unicorns-wearable-tiaras-and-unruly-robots-at-brighton-mini-maker-faire-2015/>

⁵⁴⁰ <http://journals.rcni.com/doi/pdfplus/10.7748/ns.29.16.63.s48>

⁵⁴¹ <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/contractsthatdonotguaranteeaminimumnumberofhours/mar2017>

service as part of a low wage economy and being rewarded only with internet services. They spend an enormous amount of time – in some cases statistics record that the average person spends up to two days on Facebook a week – they supply their personal data, information on their movements and behaviour that is used to program AI systems and receive nothing for it other than the use of the system that is milking them. Attempts by those users to complain or to assert their rights have proved to be often very difficult to achieve⁵⁴². Not only are the people using internet services the new oil for them in terms of revenue, they become a lubricant for its further development.

24. People whose right to receive state benefits, insurance claims or compensation, or to become British citizens, is being assessed by AI using parameters that are not based in human nature nor in natural justice. Usually these rules are designed to minimise the number of claimants and the amount paid to each one. They can produce unjust and inhuman results that cannot easily be challenged.⁵⁴³

25. Many people cannot afford to buy and maintain connected devices, and resort to pawning them at Cash Convertors or similar High Street outlets or selling them at car boot sales. Note: many people fall into all of the above categories. The UK Office of National Statistics shows that 4.6 million people, or almost eight percent of the population, are in persistent poverty.⁵⁴⁴

How can potential disparities be mitigated?

26. UBI (Universal Basic Income) was recommended by socialist MEPs in the Parliament's 2017 debate on robotics, championed by Mady Sehres Dalvaux MEP. Finland, the Netherlands, Native American communities in the USA, Uganda, Kenya, Namibia and parts of India and Latin America are running pilot schemes.⁵⁴⁵

In the United States the Roosevelt Institute has just published research⁵⁴⁶ that claims the US economy would grow by \$2.5 trillion by 2025 if every adult received an unconditional payment of \$1,000 dollars per month. And Facebook CEO Mark Zuckerberg used a speech at Harvard to voice his support for UBI⁵⁴⁷

Public perception: Should efforts be made to improve the public's

⁵⁴² <http://www.reuters.com/article/us-eu-ireland-privacy-schrems/max-schrems-the-law-student-who-took-on-facebook-idUSKCN0S124020151007>

⁵⁴³ <http://www.hartlepoolmail.co.uk/news/hartlepool-man-found-dead-on-beach-after-sickness-benefits-stopped-1-8708019>

⁵⁴⁴ <https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/articles/persistentpovertyintheukandeu/2015>

⁵⁴⁵ https://en.wikipedia.org/wiki/Basic_income_pilots

⁵⁴⁶ <https://www.cnbc.com/2017/08/31/1000-per-month-cash-handout-would-grow-the-economy-by-2-point-5-trillion.html>

⁵⁴⁷ <https://www.cnbc.com/2017/05/25/mark-zuckerberg-on-success-billionaires-should-pay-you-fail.html>

understanding of, and engagement with, artificial intelligence? Yes. If so, how?

27. By watermarking and clearly labelling interactive services that are provided by AI, such as chatbots and call centre service calls. For example, on the Marks and Spencer website and in the software developers' online community Slack the AI says "I'm just a bot but I'll try to find the answers for you". Following EU legislation⁵⁴⁸ we are now constantly alerted to the fact that websites use cookies to store our searches. This is a useful precedent for explaining AI to users and giving them the option to consciously reject or embrace it, every time. The CAPTCHA codes that require users to tick a box stating "I am not a robot" are a further step towards greater transparency: all we need is a system that requires AI to declare: "I am a bot."

In some cases, users may prefer to use a bot because it is quicker and less emotionally draining than chatting to a human. Some things are easier to say to a machine, for example giving answers to intimate health care questions. Consumers should be offered the choice of conversing with a human or not, and the process must be made honest and transparent.

6. What are the key industrial sectors that stand to benefit from artificial intelligence?

28. Robotics, health care, for example diagnostic screening such as Tumour Trace⁵⁴⁹, health administration, for example Florence by Nuance⁵⁵⁰ logistics and transportation, technology companies, farming and food production, space and defence. Small-scale and hand-made goods produced and sold locally, such as "artisanal" bakeries.

Which sectors do not?

The paradox of AI is that the industries that benefit are those that can automate processes so though the sectors where it can be deployed can reduce costs they do so by reducing headcount so those that do not benefit are those employed in technology, the hotel and hospitality sector, law, stock-markets, the finance sector, transport, farming, factories, medicine.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed?

29. They must change their ethos and put humans first. They put profits before people and this is no longer acceptable or sustainable because one third of their former consumers are out of work and can no longer afford to pay for their products and services. So, self-interest dictates that commercial companies must adopt corporate social responsibility. Likewise, one third of taxpayers will no longer be contributing to the British economy, so public spending cuts will be necessary and the private sector will need to play a bigger role.

⁵⁴⁸ http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm

⁵⁴⁹ <http://www.tumourtrace.com>

⁵⁵⁰ <https://www.youtube.com/watch?v=gO1iZ7bAsFE>

How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

30. Pay people for their personal data, as Peter Warren argues and Evgeny Morozov expounds in his book.⁵⁵¹

The UK's Open Data Institute is producing world-leading systems developed by Sir Tim Berners Lee.

Germany has very tight privacy and personal data protection laws, yet is powering ahead with a growing economy, thriving digital start-up ecosystem and healthy balance of payments.⁵⁵²

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

31. A digital divide will open between those who have the money and the skills to use AI and those who do not. This will inevitably cause societal tensions. See Cooley/CSR conference proceedings 25.05.2017⁵⁵³, also 'Can we make the digital world ethical?' Netopia 2014.

The solution to this has to be education and the use of AI to educate people, which is one of the few very bright promises of AI.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

32. This question is intrinsically linked with trade secrets and commercial confidentiality. To create, retain and profit from Artificial Intelligence, companies and non-profit entities such as university researchers need to establish and preserve their ownership of the intellectual property (IP) in it.

But for the purposes of transparency and regulation the black-box has to be able to lay out the methods by which a conclusion was reached. If we simply allow companies to say that a machine learning system used x algorithm to come to y conclusion and that is simply accepted then it is maths without showing the working out – such a practice would have profound implications for safety and could possibly leave a device or system uninsurable.

So, most AI applications should be allowed to black-box, for commercial reasons, in order to benefit the individual companies themselves and the British economy in general. They must also lay out the rules by which they are black-boxing and the rationale. The GDPR will insist on this.

33. Exceptions to this rule must be: life and death situations, discrimination,

⁵⁵¹https://books.google.de/books?id=H9ciBQAAQBAJ&pg=PA235&lpg=PA235&dq=Evgeny+Morozov+people+should+be+paid+for+their+personal+data&source=bl&ots=AX51KJIOYE&sig=nWrcw_4hSGwTmWwdgXwwiChxJiw&hl=en&sa=X&ved=0ahUKEwjB79fpoY3WAhWSJ1AKHaEIAD4Q6AEITTA

⁵⁵²https://www.destatis.de/EN/Publications/Specialized/InternationalData/G20/G20InFigures0000168179004.pdf?__blob=publicationFile

⁵⁵³ <http://www.futureintelligence.co.uk/events/>

automated justice, confidentiality:-

The launch codes for autonomous lethal weapons or weaponised software should not be automated. Again, in August 2017, for the second time in two years, a large number of influential scientists and thought leaders has called for a global ban on autonomous killer weapons.⁵⁵⁴

In a robotic milking parlour in Bedale, North Yorkshire, dairy farmer Tim Gibson has installed an AI that detects when a cow is unwell or yielding insufficient milk. It locks the gate so that she cannot return to the field. The next stop for her could be the slaughterhouse. That same algorithm could be used to evaluate vital signs in humans, in hospitals, care homes and hospices. How do we know it is not already being used, for example to support the Liverpool Care Pathway? We should be told, in this and in all other cases where lives are at stake.

Offender profiling: the public needs to know that suspects are not being targeted because of their skin colour, religion, age, lifestyle or geo-location. This should extend to systems such as facial recognition.

Automated justice, for example AIs deciding the tariff for punishment e.g. on the spot fines for parking.

AIs handling patient data in the National Health Service and private health and social care providers – patients need to know for sure that their data is not being sold on to insurance companies, for example, or decisions are being made by AI that are not transparent to the patient and his or her family.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

There is a tendency for politicians to hold up their hands and claim any legislation will stifle innovation and risk penalising domestic industries in a competitive world. The opposite is true, in the current technological age technology is making significant inroads into the lives of people and people need to have their fundamental rights maintained in the face of those incursions.

Countries and state groupings have proved to be remarkably successful in establishing the rights of individuals over those of companies and this is a process that needs to continue. People must be put firmly at the heart of the relationship with technology. Technology should serve humanity.

11. What lessons can be learnt from other countries or international organisations

32. e.g. the European Union

<http://www.futureintelligence.co.uk/2017/02/meps-vote-to-control-robots-and-put-humans-first/>

The above are our interviews with Mary Honeyball and Mady Dalvaux who spell out the need to make technology work for people and not the other way around. The EU has taken a lead in technology and has shown itself to be both fearless

⁵⁵⁴ <https://futureoflife.org/open-letter-autonomous-weapons/>

and to be prepared to take on large corporations, as shown by the penalties it has imposed on US high-tech companies. It has put in place well-received legislation on cyber security, privacy and data protection (which has become the model for the Government's recently announced privacy legislation). The EU is now planning further legislation on AI.

The World Economic Forum

34. The WEF's survey seems to fly in the face of an informed position that has been presented by Oxford University, The Bank of England, the University of Southern California (USC) and the accountants PwC. All those institutions predict job losses in excess of one third – the USC, predicts higher than that figure. The WEF draws on a global survey pool of 31,000 to draw conclusions on a population of 7.5 bn. In the case of China, for example, with its population of 1.379bn 739 people were canvassed. While the exact demographic is not shown explicitly it would appear to be relatively well-educated people with access to the researchers as opposed to goat herders in places like Africa, India or China who would have been harder to canvass and presumably may have little access to technology. Though the young people canvassed were optimistic that AI would produce jobs this is against a backdrop of young people being in worse economic circumstances than previous generations. Many AI experts admit they are not sure where the new jobs will come from.

12 September 2017

Future of Humanity Institute – Written evidence (AIC0103)

1. Background on authors

The Future of Humanity Institute at the University of Oxford researches both the technical and the societal dimensions of advanced artificial intelligence (AI) and other issues that bear on the prospects for humanity's future. Our staff brings together computer scientists, philosophers, mathematicians, social scientists, lawyers, and engineers to shed light on these issues. The founder and Director of the institute, Nick Bostrom, is the author of the best-selling 2014 book *Superintelligence*, which played a key role in fostering recent discussions around the long-term trajectory of AI. In 2016, our Institute launched the Strategic Artificial Intelligence Research Centre to analyze the policy implications of AI and develop recommendations for governments and industry. We are pleased to have the opportunity to share the perspective of our institute on several topics you are investigating.

2. Uncertainty related to the nature and timing of AI progress

2.1. Relevant committee questions: "What is the current state of artificial intelligence and what factors have contributed to this?"; "How is it likely to develop over the next 5, 10 and 20 years?"; "What factors, technical or societal, will accelerate or hinder this development?"; "Is the current level of excitement which surrounds artificial intelligence warranted?"; "What role should the Government take in the development and use of artificial intelligence in the United Kingdom?"; "Should artificial intelligence be regulated?"; "If so, how?"

2.2. Comments:

2.2.1. Artificial intelligence is currently experiencing a period of rapid and exciting progress, fuelled by several factors. Many new researchers are moving into the field, computing hardware has advanced significantly, data is more abundant than it was before, algorithmic innovations have been developed, and open source frameworks enable quicker replication of new ideas. More generally, AI research is benefiting from substantial public and private funding in various countries, with private funding becoming more dominant in the past five years than was the case previously. According to one estimate, global AI funding from tech companies, venture capitalists, and private equity firms was approximately £20 to £30 billion in 2016 (Bughin et al., 2017).

2.2.2. From the 1960s to the 1990s, AI progress often failed to live up to lofty forecasts, but more recently the opposite has taken place: across a range of benchmarks including computer performance at the game of Go and image recognition, even AI researchers have been surprised by the pace of various developments. For example, on the ImageNet challenge (one measure of AI visual capabilities, in which images are classified into 1,000 different categories), the error rate of the best

systems has dropped over the past several years from about 25% to well under 5% (the performance attained by a human). One way of characterizing recent progress is that “low level” cognitive tasks once considered recalcitrant to progress (such as visual perception) are now yielding to recent approaches in machine learning, and are now being combined with “high level” approaches such as symbolic reasoning that have seen more success in the past, bringing us closer to more general purpose, integrated AI systems.

2.2.3. Looking to the longer term, there is much uncertainty about the pace of AI progress. In the most authoritative survey of AI expert opinion to date, conducted by researchers at the Future of Humanity Institute, the non-profit AI Impacts, and Yale University (Grace et al., 2017), opinions on the future of AI varied widely among 352 experts. Human-level AI, defined here as when AI systems are better than all humans at all tasks, is a policy relevant event as it will likely be associated with radical transformations in society, the economy, and other domains. The sampled AI researchers revealed a striking consensus about our uncertainty regarding when human-level AI will be developed. The aggregate view⁵⁵⁵ of AI researchers gave a 25% chance of human-level AI in as soon as 20 years, but also a 25% chance that it won't arrive for 100 years. Even after adding additional uncertainty to these estimates, the policy-relevant conclusion from these assessments is that we should not base our policy plans on any particular timeline for human-level AI: it may take many decades, but it may also arrive much sooner than we realize.

2.2.4. The surveyed researchers also revealed a nuanced perspective on the potential consequences of human-level AI, with the median respondent giving a 45% probability of a “good” or “extremely good” outcome for humanity, but also 10% and 5% probabilities of a “bad” or “extremely bad (e.g., human extinction)” outcomes. For discussion of these risks, see (Bostrom, 2014).

2.3. Recommendations:

2.3.1. In light of the range of expert opinions on future AI developments, and the benefits of preparedness, we recommend that the UK government not make strong assumptions about how quickly AI will develop. On issues such as technological displacement in the labor market in the coming decades, underestimating and under-preparing for AI's impact could result in major societal disruption and lost opportunities for shared economic and social gains. For example, the aforementioned survey by Grace et al. (2017) found that experts foresee many jobs being susceptible to automation in the next two decades, such as retail jobs and truck driving.

⁵⁵⁵The mean of the individual cumulative distribution function estimates, also called the “mixture” distribution.” The median is similar.

2.3.2. However, we also note that few researchers think it likely that human-level AI will be developed in the very near future (less than ten years), and we recommend not taking substantial action motivated by these concerns until more robust policy proposals for how to best navigate this transition have been proposed and vetted. For discussion of desiderata for such policy proposals, see a recent working paper by [Bostrom, Dafoe, and Flynn \(2017\)](#).

3. Near term challenges associated with AI

3.1. Relevant committee questions: “How can the general public best be prepared for more widespread use of artificial intelligence?”; “What are the ethical implications of the development and use of artificial intelligence?” “How can any negative implications be resolved?”; “What role should the Government take in the development and use of artificial intelligence in the United Kingdom?”

3.2. Comments:

3.2.1. One area in which AI is likely to have a significant impact in the near term is on the nature of work. Experts vary on the speed with which job displacement related to AI might occur, and the extent and nature of jobs that will be created as a result of AI. But it is widely believed in the AI community that over the next few decades, large impacts are likely, and our survey discussed above suggests high confidence that some jobs in retail will be susceptible to automation.

3.2.2. AI is likely to generate myriad other social and economic challenges. For example, there are legitimate and challenging political and legal issues pertaining to the appropriate development and use of autonomous vehicles, the acquisition, use, and ownership of people’s data, and the use of AI in important decision making contexts such as the granting of loans and parole.

3.2.3. AI is likely to have potent security implications, including beneficial applications such as more effective cyber-defenses, as well as myriad possible malicious uses by terrorists and criminals. Some novel forms of attack made possible by AI, such as large-scale, highly effective automated “spear phishing” and delivery of lethal force by repurposed consumer drones, are troubling. The government will need foresight to realize these positive applications of AI to security and prevent or mitigate the consequences of the negative applications. We outline these concerns in a forthcoming public report, based on a February 2017 workshop.

3.3. Recommendations:

3.3.1. We recommend the UK government prepare for the possibility of significant job displacement, as well as creation, as a result of the deployment of AI in the coming decades. We recommend that the UK government consult with (among others) experts at the University of Oxford such as Michael Osborne and Carl Frey who have done seminal work on this topic, and reevaluate education and job retraining

programs in light of expert views on the future of AI-related job displacement.

- 3.3.2.** We recommend that the UK government pursue novel, privacy-preserving data governance systems to ensure that the benefits of AI in health research, security, and other areas are realized while also ensuring appropriate protections of individual data. Ongoing work in the research area of secure and private machine learning (Papernot and McDaniel et al., 2016), for example, is potentially useful as the UK government seeks to be a leader in spurring AI innovation while protecting important societal values. Likewise, in the area of crime and terrorism prevention, AI has the potential to be a boon for security, and the UK can lead the way in developing innovative approaches to privacy-preserving AI-augmented surveillance.
- 3.3.3.** We recommend that the UK government consider the risks of AI being used for harmful purposes by state and non-state actors and take steps to better understand and reduce such risks. For example, some promising interventions would be “red team” exercises to determine the threats to government systems, analysis of lessons learned from other dual use technologies such as biotechnology, and exploration of the legal implications of AI-enabled threats (for example, how “data poisoning” attacks aimed at machine learning systems might be treated under existing or future laws).
- 3.3.4.** Furthermore, given the infrequency with which best practices in cybersecurity are adopted by individuals and organizations, we suggest that the UK government to consider recent proofs of concept of offensive AI applications in cybersecurity as a “wake up call” regarding the pace of innovation in this space, and as a reason to increase its commitment to the promotion of cybersecurity best practices.
- 4. Long term challenges: building AI for the common good**
- 4.1. Relevant committee questions:** “What are the ethical implications of the development and use of artificial intelligence?” “How can any negative implications be resolved?”; “What role should the Government take in the development and use of artificial intelligence in the United Kingdom?”
- 4.2. Comments:**
- 4.2.1.** Over the long term, AI is likely to exceed human performance in most cognitive domains. This poses substantial safety risks, described in detail in (Bostrom, (2014) and Amodei and Olah et al. (2016) and endorsed as worthy of study by thousands of AI researchers (Future of Life Institute, [2015](#), [2017](#)). One challenge, among others, is to ensure that the (implicit) goals of extremely competent AI systems are precisely what, upon reflection, humans would want for them. This challenge was foreseen by some pioneers of AI and cybernetics, such as Norbert Wiener who said in 1960: “We had better be quite sure that the purpose put into the machine is the purpose which we really desire.”

- 4.2.2.** Active research on AI safety is being conducted by labs in industry (including DeepMind in London), non-profits (such as OpenAI in the United States), and in academia (including at UC Berkeley, the University of Montreal, and the Future of Humanity Institute in Oxford). While these problems seem solvable in principle—we are not aware of any reason why an arbitrarily intelligent AI system, appropriately designed, could not be aligned with human values--in practice addressing this issue seems likely to require substantial research, foresight, and prudence.
- 4.2.3.** In the coming decades, AI developers will face a variety of incentives and pressures. Scientific, economic, and other forms of competition, especially between countries, could lead to substantial pressure to quickly develop and deploy advanced AI systems. These pressures risk leading to insufficient attention to safety and other social considerations. We will be better off if leading AI developers in all countries commit to, and are able to work towards, developing AI for the common good.
- 4.2.4.** If designed and governed appropriately, AI has the potential to be extremely positive-sum in its societal impacts--for example, it may enable rapid economic growth and improved health. Ensuring these benefits are realized is an additional reason to pursue cooperative development of AI and avoid potentially dangerous racing.
- 4.3. Recommendations:**
- 4.3.1.** We recommend that the UK government step into a global leadership role in developing international norms and institutions for building AI for the common good: in a way beneficial for humanity as a whole. This *common good principle* was articulated by Bostrom (2017), endorsed by many signatories in the AI community in the Asilomar AI Principles (Future of Life Institute, 2017), and discussed further in a report from the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (2016). The UK government could begin by making a commitment to fostering AI research and development for the common good. Such a commitment would signal the UK's leadership in AI governance, commensurate with its prominent role in AI and AI safety research. What specifically this commitment should entail, and how best to realize it, will require creative exploration in partnership with industry, researchers, the public, and other countries. A key institution to collaborate with on this front is the Partnership on Artificial Intelligence to Benefit People and Society, which includes many relevant companies as well as non-profits, such as the Future of Humanity Institute, as partners.
- 4.3.2.** We additionally recommend that the UK government explore the possibility of creating or joining international research efforts in the domain of AI R&D, and doing so in part on the basis of which projects are committed to the common good. The UK government would thereby contribute to building beneficial norms and institutions

promoting international cooperation on developing AI. The UK government should support existing efforts for international dialogue and governance about AI, such as that being promoted by the United Nations.

5. Research areas appropriate for public support

5.1. Relevant committee questions: “In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable?”; “When should it not be permissible?”; “How can the general public best be prepared for more widespread use of artificial intelligence?”; “What are the ethical implications of the development and use of artificial intelligence?” “How can any negative implications be resolved?”; “What role should the Government take in the development and use of artificial intelligence in the United Kingdom?”

5.2. Comments:

5.2.1. The UK is in a leading position today in the field of AI, and has an opportunity to build on this lead with a long-term commitment to AI research and development.

5.2.2. While AI safety and policy research are currently being conducted, in part, by private actors, it is probable that the level invested is socially suboptimal given that this area is characterized by substantial global positive externalities. Relatedly, some AI applications that are unlikely to be immediately profitable (such as applications specifically aimed at achieving the UN Sustainable Development Goals) are opportunities for public investment to ensure that these public goods are created. Finally, research aimed at better understanding the policy implications of AI, while again being supported in part by private actors, is clearly in the broad public interest and may be currently undersupplied.

5.3. Recommendations:

5.3.1. The UK government should double down on its strong competitive position in AI by investing substantially in AI research, development, and education. The UK government should also develop a robust, long-term funding portfolio supporting research on AI safety, policy, and socially beneficial applications. Examples of technical safety research includes work building AI systems that cooperatively learn human preferences and social norms (Hadfield-Menell et al., 2016; Christiano et al., 2017) and designing AI systems that are reliable even under adversarial attack (Papernot and McDaniel et al., 2016; Amodei and Olah et al., 2016). Examples of policy research include characterizing the potential global externalities from AI development and crafting institutions for international cooperation. These investments should be informed by ongoing dialogue with experts in AI and other areas in order to ensure that new government funding complements existing research trajectories in industry and academia.

6. Other recommended government actions

- 6.1. Relevant committee questions:** “What role should the Government take in the development and use of artificial intelligence in the United Kingdom?”
- 6.2. Comments:**
- 6.2.1.** There is an ongoing flow of talented AI researchers from academia into industry, and as a result of demand exceeding supply, these researchers can currently command very high salaries. These salaries as well as other benefits of working in industry (such as proximity to other talented researchers and access to large amounts of data and computing power) present a formidable obstacle to the UK government (and academia) in recruiting AI experts, especially in the area of machine learning.
- 6.2.2.** At the same time, as AI is increasingly adopted in society, it is perhaps more important than ever before that the UK government recruit such experts, suggesting a need for creative thinking.
- 6.3. Recommendations:**
- 6.3.1.** We recommend that the UK government consider creative approaches for recruiting AI experts (including both technical and policy experts) into government, in order to put the government in a better position to proactively address problems and exploit opportunities as they arise. We recommend that the government consider lessons learned from other domains, such as finance and law, where competition for talent with the private sector has been fierce, and consider novel initiatives such as special authority for a department to pay higher than usual salaries.
- 6.3.2.** Finally, we note that beyond just salaries, it will be important to motivate recruits with an exciting mission (Brundage and Bryson, 2016). The formation of a new agency for the purpose of developing and funding socially beneficial AI, and steering AI’s social impacts in a positive direction, might be one such approach. A standing Commission on Artificial Intelligence, as suggested in written evidence submitted by the Future of Humanity Institute and others previously and endorsed by the Science and Technology Committee’s report on robotics and intelligence, could be a focal point for recruitment.
- 7. Offer of further dialogue:** We welcome the opportunity to provide further information. The Future of Humanity Institute is particularly well placed to help the Lords Select Committee understand issues related to long-term AI safety, the range of expert opinions on AI’s future, the near-term intersection of AI and security, international dynamics around AI, and the global governance of AI.

Miles Brundage and Allan Dafoe
On behalf of the Future of Humanity Institute
University of Oxford
<https://www.fhi.ox.ac.uk>

References

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. 2016. "Concrete Problems in AI Safety," available online at <https://arxiv.org/abs/1606.06565>

Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press: Oxford.

Bostrom, N. 2017. "Strategic Implications of Openness in AI Development," *Global Policy*, Vol. 8, Issue 2, May 2017, pages 135-148, available online at <http://onlinelibrary.wiley.com/doi/10.1111/1758-5899.12403/full>

Bostrom, N., Dafoe, A., and Flynn, C. 2016. "Policy Desiderata in the Development of Machine Superintelligence," working paper, Future of Humanity Institute, available online at <https://nickbostrom.com/papers/aipolicy.pdf>

Brundage, M. and Bryson, J. 2016. "Smart Policies for Artificial Intelligence," available online at <https://arxiv.org/abs/1608.08196>

Bughin, J, Hazan, H., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., and Trench, M. 2017. "Artificial Intelligence: The Next Digital Frontier?," McKinsey Global Institute discussion paper, available online at <http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/how-artificial-intelligence-can-deliver-real-value-to-companies>

Christiano, P., Leike, J., Brown, T., Martic, M., Legg, S., and Amodei, D. 2017. "Deep Reinforcement Learning from Human Preferences," available online at <https://arxiv.org/abs/1706.03741>

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., and Evans, O. 2017. "When Will AI Exceed Human Performance? Evidence from AI Experts," available online at <https://arxiv.org/abs/1705.08807>

Future of Life Institute, 2015. "An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence," text and signatories available online at <https://futureoflife.org/ai-open-letter/>

Future of Life Institute, 2017. "Asilomar AI Principles," text and signatories available online at <https://futureoflife.org/ai-principles/>

Hadfield-Menell, D., Dragan, A., Abbeel, P., and Russell, S. 2016. "Cooperative Inverse Reinforcement Learning," available online at <https://arxiv.org/abs/1606.03137>

Papernot, N., McDaniel, P., Sinha, A., and Wellman, M. 2016. "Towards the Science of Security and Privacy in Machine Learning," available online at <https://arxiv.org/abs/1611.03814>

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, 2016. *Ethically Aligned Design: A Vision for Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*, Version 1. IEEE. Available online at http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

5 September 2017

Dr Samantha Gallivan – Written evidence (AIC0185)

Written Evidence for Select Committee on Artificial Intelligence.

Author:

Samantha Gallivan MBBS BSc. FRCS (Tr. & Orth).

I am an Orthopaedic Surgeon, and work as a Post CCT Fellow at St George's Hospital, Tooting. I entered the GMC specialist register in 2016 and am a Fellow of the Royal College of Surgeons of England. This year, I joined the Royal Society/ Leverhulme Centre for the Future of Intelligence AI Narratives workshops and visited Columbia University as a Winston Churchill Memorial Trust Fellow. I have experience of coding iOS apps and am an M Ed. candidate at Imperial College London, where I obtained my primary medical degree. I have current experience in data driven healthcare policy research in the 'Getting It Right First Time' project, investigating the causes of variation in the quality of NHS healthcare.

Artificial Intelligence & Healthcare.

This submission aims to answer four of the questions posed in the call for evidence (in **bold**). All answers relate to the impact of AI on healthcare and the NHS, and the final section proposes some solutions to the challenges ahead. This submission is personal evidence and does not reflect the opinions of my employers nor academic institutions with which I have research links.

Working Definition

'Artificial Intelligence' (AI) refers to the ability of software and robotic systems to gather and use information to help make decisions. It is difficult to create a precise definition of intelligence, both in humans and artificial systems, but most researchers agree these abilities are required:

- Sensing
- Reasoning
- Knowledge representation- creating ontologies to make sense of the world
- Planning
- Learning
- Natural Language Communication

Some researchers believe that artificial general intelligence is possible, whereby a machine could perform all of these intellectual tasks by taking problem solving skills that were learned in one domain and applying these to another. Current applications of AI tend to be 'narrow' in which non-sentient intelligence is

concentrated on solving a well-defined problem. Familiar examples of these successful AIs include email filtering, mobile map applications and the manipulation of social network news feeds.

The Pace of Technological Change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

2. Is the current level of excitement which surrounds artificial intelligence warranted?

There have been attempts to build sophisticated computer decision support tools in medicine since the 1990s^{556, 557}. These were designed to mimic clinical reasoning and so were modelled on the decisions made by an expert panel of doctors when faced with a medical question. These early tools can be imagined as decision trees. New decisions are represented by branches in the tree and these can terminate, branch further or feedback to a higher level. The shape of the tree is dictated by the software designer who weights the likelihood of each diagnosis based on medical evidence from trials and the input of clinical experts. The weightings of these decisions should not be thought of as a static 'on/off' switches. It is possible to mathematically model how likely a particular outcome is in the context of our prior knowledge about the system and how these probabilities might change with time. The intelligence in this type of technology lies in the ability to access and weigh up huge volumes of information faster than a human physician. Unfortunately, these models tend to be inflexible and struggle with the nuances of real life medicine, requiring too much clinical time to be useful and are also vulnerable to any biases inadvertently 'baked in' by the designing team. Modern healthcare AI systems seek to learn from the bottom up by drawing inferences directly from the data itself, particularly in fields such as radiology and histopathology where vast amounts of visual information are available⁵⁵⁸. As doctors, we are trained to 'read' x Rays in a systematic manner, guided by our knowledge of medicine learned outside the context of the image. We look for the subtle line of fluid crossing a chest x ray of a patient with heart failure or white seeding in the lungs in tuberculosis, and are able to conjure these disease processes in our imagination. By contrast, a machine learning

⁵⁵⁶ Features of effective computerised clinical decisions: metaregression of 162 randomised trials. Roshanov, PS, Fernandes, N. et al BMJ 2013 Feb 14; 346

⁵⁵⁷ Decision tools in health care: focus on the problem, not the solution. Liu, J, Wyatt, J & Altman, D. BMC Med Inform Decis Mak. 2006; 6:4

⁵⁵⁸ Automated analysis of retinal imaging using machine learning techniques for computer vision. De Fauw, J Keane, P, Tomasev, N et al. F1000 Res. 2016; 5:1573.

system does not 'know' these as disease states but is trained to recognise patterns of abnormality. These may be through hand designed features, where data labelled by experts (such as radiologists) helps train the system or through unsupervised learning which automatically learns feature representations through massive deep learning algorithms. For these to succeed, huge volumes of medical images are required, ideally already annotated by experts⁵⁵⁹, hence the interest of global AI companies in NHS data.

It seems likely that in the next 5-10 years, basic medical science, medical imaging and pathology will be revolutionised by machine learning. Greater automation of image processing should help reduce errors through fatigue or inattention and free up radiologists and pathologists to concentrate on cases that pose true diagnostic dilemmas requiring clinical wisdom to solve. It is exciting to speculate that deep learning algorithms might even 'see' hidden correlations between protein shapes or imaging data that helps hone the design of new drugs and therapies in an unexpected way that changes how we think about disease progression.

In addition to imaging and pathology, some health AI start-ups claim to have made progress in triage with the development of chatbots that simulate human conversation. Earlier this year, Babylon (a London based company) trialled an app in North Central London to 1.2M NHS patients, that is designed to offer 'symptom advice' through a text conversation. As we have seen previously, any machine learning algorithm is dependent on large volumes of high quality training data and particularly in the case of a decision support tool, requires deep content knowledge at the design stage. Interestingly, Babylon have partly trained their algorithm on health data from their internal database of paying users, who one might assume represent a thin slice of tech savvy, wealthy patients⁵⁶⁰. It would be informative to see how their app performs in a well-designed clinical trial exploring both clinical outcomes and the experience of using the app, with a more diverse, representative patient population. Some concerns have been raised in recent weeks as to how AI apps such as these are tested and this now appears to be a critical area for research⁵⁶¹.

⁵⁵⁹ Applying machine learning to automated segmentation of head and neck tumours and organs at risk on radiotherapy planning CT and MRI scans. Chu, C, De Fauw, J, Tomasev, N et al. F1000 Res. 2016; 5:2104

⁵⁶⁰ Public Talk- Wilhelm van Der Walt, python tech lead- Python Meetup- Lessons Learnt Building a Chatbot in Python, babylon offices, June 29th 2017

⁵⁶¹ <https://doi.org/10.1136/bmj.j3980> Innovation without sufficient evidence is a disservice to all. McCartney, M. BMJ 2017; 358:j3980

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

How should AI enabled apps and algorithms be tested in healthcare?

It must not be forgotten in the enthusiasm for AI in medicine, that every medical decision is socially situated and interacts with other decisions in a dynamic, complex environment. In order to test a medical algorithm, the impact on the whole system needs to be assessed in a real hospital setting, not simply in simulation in a computer lab. Healthcare decisions are unlike other transactions; they are laden with ethical implications and this must be accounted for in any algorithm design. Machine learning works by training classifiers to define boundaries and identify clusters in data. Doctors also recognise patterns in illnesses but we work within our society and so are part of the interplay of norms and meaning that structure the expectations we have for medical professionals.

Safety- past lessons from industry, regulation and research.

By contrast to the 'move fast and break things' attitude of Silicon Valley with its implication of ongoing iterations of improvement, medicine has learned that even those with the best of intentions can create permanent, yet inadvertent harm. Sometimes these negative effects occur through mechanisms unimagined at the time of device or drug development and so only come to light through scientific testing. Pharmaceutical regulators across the world prevent poorly tested drugs from being released and enforce diligent post marketing surveillance. An early success of the FDA was to block the sale of Thalidomide in the US when an FDA researcher, Dr Frances Kelsey had concerns about a single case study of reported nervous system side effects⁵⁶².

Independent academic research is also vital to ensure that interventions that appear beneficial, do not in fact cause unexpected harms. In 1989-91, the CAST trial tested the hypothesis that anti-arrhythmic drugs would lower mortality after a heart attack by suppressing abnormal heart rhythms⁵⁶³. To the surprise of the cardiology community, mortality was increased by use of these drugs and evidence from this trial revolutionised the treatment of heart disease.

Without rigorous formal testing, it is impossible to know whether the claims made by tech companies as to the benefits of their algorithms are true and impossible to prevent unnecessary harm. We hold pharmaceutical companies responsible for the safety of the drugs they develop within a strict regulatory

⁵⁶² https://www.nytimes.com/2015/08/08/science/frances-oldham-kelsey-fda-doctor-who-exposed-danger-of-thalidomide-dies-at-101.html?_r=0 New York Times Obituary of Dr Frances Oldham Kelsey, August 7th 2015

⁵⁶³ Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. CAST Trial Investigators. NEJM. 1989 August 10;312 (6)

environment. It seems regrettable that we do not hold technology companies to the same standard. Introducing an app that alerts for one disease such as acute kidney injury may help patients with that condition, but what of those patients with no kidney disease who have had care and attention diverted away? Who decides which conditions 'deserve' an alert in an imagined hierarchy of disease and how does the algorithm decide which alert takes precedence? It has been shown repeatedly that gender and racist biases permeate AI algorithms as a reflection of the flawed society in which we live⁵⁶⁴; why do we assume that healthcare AI will be any more immune to this effect?

At the heart of many technology companies' assumptions about healthcare, appears the positivist belief that if only more data points could be harnessed and analysed, then the 'truths' of medicine would be revealed. While it is true that pathology, imaging and population medicine may benefit from machine learning, it is a false assumption to believe that clinical decision making can be tackled in the same fashion. Medical Artificial Intelligence that combines clinical wisdom, ethical decision making and bias free algorithm design is a distant dream.

Who benefits from the sale of NHS data?

Some commentators have likened the recent rush for large public data sets to the oil booms of the nineteenth century, on which the world's economy still thrives⁵⁶⁵. The Earth's oil seemed then to be an endless natural resource, whose value was silently created by the slow transformation under pressure of the corpses of unknowing creatures that died millions of years ago. The annotated medical data of the NHS is indeed our modern oil, but it is formed by the goodwill and skill of the patients and staff who co-created it. This data collects at every hospital appointment, every blood test, phone call and operation. It is made by the porters who bring a patient down to the scanner with a joke to relax them, the expertise of the radiographer whose skill obtains a perfect image, the radiologist who methodically annotates and reports on the scan and above all, the patient, whose body is now represented as pixels on a screen. We share our data with the medical team treating us on the understanding it will directly help our care and with further consent, we share our data to help others in the future through medical research. The NHS data set is a resource that is built on the care of tens of millions of people, covering every medical and surgical specialty and united by a unique NHS number that enables linking across data sets. Recent deals with private companies such as Google DeepMind have seen large data sets leave the NHS in cooperative research projects. Unfortunately, it is still unclear as to whether the IP associated with any algorithm developed in these deals will ever return a financial benefit to the NHS. Quite reasonably, these international companies argue that they are the forefront of innovation and point to the investment they have made in bringing new technologies to the NHS.

⁵⁶⁴ Inspecting Algorithms for Bias. Spielkamp, M, MIT Technology Review, June 12th 2017.

⁵⁶⁵ 'The World's most valuable resource is no longer oil, but data' The Economist- Leader. May 6th 2017

These technologies are impossible without clinical data and we should not be held hostage to release it for free.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

1) AI Healthcare Regulation.

Oversight for different aspects of technology in healthcare currently falls to the MHRA (Medicines and Healthcare products Regulatory Agency), the ICO (Information Commissioner's Office) and the CQC (Care Quality Commission). I suggest a single regulator that draws on expertise from all three but concentrates solely on technologies that incorporate automation, artificial intelligence and data sharing with external technology companies in healthcare.

AI poses a particular challenge for existing regulators, because as technologies evolve and users alter the way in which they interact with the device, the classification of apps and software changes. We need much clearer guidance on modern decision support apps- what is the legal implication of a mistake made by the app without the raw data being available to the clinician? How do we audit the outcomes of apps and AI enabled software to ensure they are safe and unbiased? What steps can be taken to interrogate the decision making of AI algorithms, even when their design is inherently opaque? This will require co-operation with software designers and their international parent companies. I suggest a summit to bring together the leaders in the field to help agree some of these standards and encourage new research into AI technologies that are more easily comprehensible to humans⁵⁶⁶.

We must also clarify the roles of Data Controller and Data Processor as defined in the Data Protection Act in light of recent controversies concerning The Royal Free Hospital and Google DeepMind⁵⁶⁷.

There is a culture clash between the hyperbole and possibly unfounded hype from technology companies and the over-reticence of the medical profession to engage in new practices. A middle way might be NHS sponsored trials where technology companies tender for access to data in a transparent and accountable manner. With appropriate patient consent, NHS clinical data is then shared with partners to develop applications with strict oversight in accordance with the

⁵⁶⁶ Towards Deep Symbolic Reinforcement Learning. Garnel, M, Arulkumaran, K, Shanahan, M. <https://arxiv.org/abs/1609.05518>

⁵⁶⁷ Google DeepMind and healthcare in an age of algorithms. Powles, J, Hodson, H. Health and Technology. 16th March 2017 <https://link.springer.com/article/10.1007/s12553-017-0179-1>

Caldicott Principles. These apps are then trialled in a controlled environment in the NHS and the data assessed by an independent research panel alongside the developers with successful apps gaining UK regulatory approval. The ROI for the NHS investment in trials and regulation is held in an IP share of the algorithm when it is sold on by the parent company. At present, there is little indication that commercial algorithms built on NHS data will financially benefit the UK and little assurance that they are effective and safe. If today's healthcare data is yesterday's oil, then the NHS is potentially giving away tomorrow's sovereign wealth fund in poorly designed data sharing deals.

6 September 2018

Ms Joanna Goodman, Dr Paresh Kathrani, Dr Steven Cranfield, Chrissie Lightfoot and Michael Butterworth – Written evidence (AIC0104)

Artificial Intelligence in Legal Services: the boundary between disruption and evolution

Summary

This submission covers artificial intelligence (AI) currently being used to deliver legal services and broaden access to legal advice, the ethical issues that lie at the boundary of its disruption and evolution, and the implications for legal education.

The implications for legal practice:

AI is changing our world, and the legal sector carries particular responsibility. The Select Committee on Artificial Intelligence recognises that regulation is needed to make sure AI is used positively: to improve access to legal services and develop new ones, increase access to justice, and make legal processes fairer, rather than more rigid, and deliver and enhance legal education. The challenge is to ensure that regulation does not limit progress or tech innovation. The legal sector is crucial to this as it is responsible for delivering legislative promises, and protecting the rights of businesses and individuals – and this means identifying how AI can help – its potential and its limitations.

The implications for legal education:

The augmented (by intelligent automation and online services) lawyers of the future will not only have to be able to understand and work with the outputs of AI technology that (helps to) deliver legal services, and legal decisions, including online courts; they will also have to consider the legal and ethical implications raised by the AI-powered systems, applications and devices routinely used by businesses and individuals (from smart phones to smart homes, autonomous and semi-autonomous vehicles), and how these technologies, which may include tracking and facial recognition, and the data they produce, affect the development, application and enforcement of the law, including decision-making at multiple levels and the rights of people and organisations.

Co-authors

The submission was initiated by Joanna Goodman, a journalist and author, who writes about emerging technology and AI for multiple publications, including *The Guardian* and *The Times* business supplements and I am IT columnist for the *Law Society Gazette*, writing a monthly feature on technology in the legal sector. She has conducted specific research into the development of AI in the legal sector. Her book, *Robots in Law: How Artificial Intelligence is Transforming Legal Services* (2016) is a practical introduction to the application of AI to legal

services and has attracted an international readership across law firms and law schools. She continues to research, write and speak on this topic.

The co-authors are members of the University of Westminster Law School's interdisciplinary and diverse Law and Technology Hub. This group was formed in 2017 by Dr Paresh Kathrani, senior lecturer in law, and includes academics, technologists, lawyers, entrepreneurs, and commentators. Its rationale was the realisation that AI's disruptive force is transforming the law and legal education will have to change dramatically.

Westminster University academics: Dr Kathrani's research includes the role and application of emerging technology in delivering legal services, including access to justice. He has hosted several events and roundtables, and co-authors on this piece have served as panellists on these. Dr Steven Cranfield, senior lecturer in leadership and professional development broadens the academic perspective into practical issues around training and developing the lawyers of the future, as well as educating today's lawyers about the challenges and opportunities AI presents for legal service delivery.

Legal AI practitioners: Chrissie Lightfoot, lawyer, legal futurist, and author of *The Naked Lawyer* series of books, has recently directed her experience and energy into a legal technology/AI start-up, Robot Lawyer LISA, a commercial legal chatbot. Michael Butterworth is a commercial technology associate at leading IT law firm Kemp Little, which advises technology companies and has used commercial AI platforms and internal developers to design and built AI applications that support the business by increasing back-office productivity as well as developing new client facing products and services.

We have focused on the Select Committee's questions in relation to the impact of AI on legal services and on legal education.

Submission

The pace of technological change

I. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1. Defining artificial intelligence as it applies to the legal sector.

The Oxford English Dictionary defines AI as: *'The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision making and translation between languages'*

PwC's more recent, detailed definition is directly relevant to the legal sector (Source: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>)

'In our broad definition, AI is a collective term for computer systems that can sense their environment, think, learn, and take action in response to what they're sensing and their objectives. Forms of AI in use today include, among others, digital assistants, chatbots and machine learning. AI works in four ways:

Automated intelligence:

Automation of manual/cognitive and routine/non-routine tasks.

Assisted intelligence:

Helping people to perform tasks faster and better.

Augmented intelligence:

Helping people to make better decisions.

Autonomous intelligence:

Automating decision making processes without human intervention.'

Law is one of the industries that are already putting these theories into practice. This is the existing state of legal AI. But its potential is far greater. The boundaries could certainly be pushed further.

- 1.2. Artificial intelligence in legal services is narrow – it is applied to specific processes. Although, there are no 'robot lawyers' yet, there are applications that carry out legal tasks. These fall into two main categories: AI platforms and applications that improve productivity for law firms, legal process outsourcers, and corporate law departments; and client/public facing AI applications, like chatbots, intelligent mobile applications (apps) and online services that help people understand their legal issues and connect them with the right advice. Leading product
- 1.3. Popular legal AI applications are:
 - M&A due diligence, where leading vendors include RAVN iManage, Kira Systems and Luminance. They handle fact checking, document/contract review and analysis, using machine learning to identify anomalies and similarities.
 - Legal research, e.g. ROSS Intelligence which conducts paralegal work, answering questions, identifying precedents, knowledge and expertise
 - Compliance, where platforms such as Neota Logic enable law firms and legal departments to develop tools for advising on specific legislative developments.

- Predicting outcomes of court cases e.g. LexisNexis Lex Machina; Premonition
 - AI-powered software including triage for managing corporate legal departments e.g. Riverview Law and back office support e.g. Thomson Reuters Elite Workspace Assistant, Helm 360's Termi practice management chatbot
 - New ways of providing/accessing legal services – Chatbots which help the public with legal issues e.g. Joshua Browder's DoNotPay which challenges parking fines, advises on housing and immigration and has recently expanded to cover other legal issues affecting the public; Lawbot, now rebranded Elexirr. Chatbots which help with commercial issues e.g. Robot Lawyer LISA.
- 1.4. Catalysts in the UK include market liberalisation, which may contribute to the UK leading the world in legal AI. Cost is a significant factor. AI's horizontal scalability helps firms gain a commercial advantage by delivering more for less while maintaining their margins and enables altruistic organisations to produce scalable services that go some way to addressing legal aid cuts.
- 1.5. Law is traditionally perceived as slow to change. However, this has not been the case for legal AI. In just two years since legal AI start-up RAVN Systems introduced Berwin Leighton Paisner's 'contract robot', LONald, an AI-powered software programme that automatically verifies property details on real estate contracts with the Land Registry, legal AI has hit the mainstream with law firms of all sizes and profiles introducing AI tools to boost productivity and develop new customer facing services. RAVN Systems was recently acquired by popular document management system vendor iManage (formerly part of Autonomy and then HP). The big legal information and software vendors, LexisNexis and Thomson Reuters both have acquired or developed AI offerings and interfaces with smart technology such as the Amazon Echo.
- 1.6. Another catalyst is the global lawtech start-up movement, championed in the UK by Jimmy Vestbirk's Legal Geek meetups. Start-ups and incubators are supported by the legal establishment, including the Law Society, magic circle and international law firms and major technology vendors. The number of lawtech start-ups and patents has seen a 484% increase in since 2012 (Source: Thomson Reuters <https://www.thomsonreuters.com/en/press-releases/2017/august/thomson-reuters-analysis-reveals-484-percent-increase-in-new-legal-services-patents-globally.html>). However, many new AI products present new approaches to the same tasks: intelligent data/contract analysis; intelligent process automation; and legal research. On the customer facing side, law firms are launching incubators and partnering with established vendors to develop AI tools that that they can white label and license to clients, and apps to advise

on legislative developments e.g. Brexit or regulatory compliance, for example, with GDPR.

- 1.7. Over the next five years, if current trends continue, it is likely that AI augmented law firms and lawyers will become the 'new normal'. Law firms will need fewer trainees and junior associates and this is likely to change law firm structure from its current pyramid shape into a diamond shape, with the majority of human lawyers at mid-level rather than the bottom. Lawyers will need to learn to use AI tools and work with the outputs of AI.
 - 1.8. Butterworth identifies some challenges that need to be addressed to enable legal AI to live up to its promises:
 - Technologists and lawyers need to work closer together, especially from a design perspective. This may have education and training consequences. Features such as rigorous audit trails, a certain level of transparency over the algorithms and good user interfaces are important. These will give the lawyers confidence in the tool, the ability to demonstrate the basis for the advice to clients and ensure that the technology fits into existing workflows.
 - Some AI tools add value by making existing processes more efficient, whereas the more impactful AI tools will be ones that make new services and access to justice possible or serve new markets (e.g. the DoNotPay bots). Development of such game-changing AI tools will take time. However, one important step to be considered now is what services and markets are capable of being transformed by AI.
 - Many AI tools use machine learning, which requires large, unbiased datasets. It is crucial to identify the right dataset and to provide continuous supervision – especially in terms of client confidentiality, which is part of the professional code of conduct. Again, technologists will need to work together to share datasets and to train and supervise the algorithm's learning.
2. Is the current level of excitement which surrounds artificial intelligence warranted?
 - 2.1. As Pathrani observes, AI is just one sub-stream in the wider technological zeitgeist. But it's a fundamental one. Perhaps it is the engine that's likely to run everything else. This is because as wider technological demand increases, there will be an even greater call to liberate people from the mass of technology around them by simplifying things to make them efficient. In other words, AI will be needed to manage the wealth of tech that people are likely to find themselves surrounded by in the next few years. We are already seeing that to some degree with virtual assistants and the internet of things – and the use of dashboards, triage systems and research tools.
 - 2.2. Notwithstanding the hype around legal AI which has catapulted legal technology into the mainstream media, the excitement is relative. It

very much depends on who we are talking about. People find AI exciting because it is new and it cuts down on a whole amount of time. But it will get to a point where it won't be exciting anymore. It will be like email – revolutionary when it first came out, but now the norm.

- 2.3. But in terms of lawyering, AI will have a big effect. It is important to recognise that legal AI isn't something that exists distinct from the legal education and professional world – but has a symbiosis with it. Law firm clients will also use AI in their homes. Hitherto, gathering and analysing evidence and information was mainly based on human interaction between lawyers and their clients. AI systems have changed the way in which information is gathered and analysed. Proformas and online inputs are replacing the need to see a lawyer in some cases – and programmes are able to spot patterns and analyse contractual terms.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?
4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?
 - 4.1. Lightfoot refers to the positive economic impact of AI, including productivity gains from businesses augmenting their labour force with AI technologies and increased consumer demand for personalised and other AI-enhanced products and services.
 - 4.2. In the legal sector, the Solicitors Regulation Authority (SRA) research, 'Improving access – tackling unmet legal needs' <http://www.sra.org.uk/risk/resources/legal-needs.page> reports that the current legal services market is inaccessible to many people who require help and advice. The two main barriers are cost and lack of consumer information.
 - 4.3. Deploying AI in the legal eco-system already negates some of these barriers and problems. Improving access to legal services through the use of smart technology has already begun and will inevitably accelerate.
 - 4.4. But the real progress to tackle the unmet need problem (the problem of the legally unrepresented, underserved and neglected) and provide legal value to disgruntled consumers and businesses that currently use human lawyers, is also coming from positive disruptors from outside the legal profession and this is likely to continue.
 - 4.5. The impact of chatbots and AI platform technologies that automate legal documents and legal advice is set to increase. For example, Robot Lawyer LISA was launched in April this year with a vision to fundamentally change the way the average person and business can

access affordable self-help, self-service, high-quality legal insight, documentation and advice.

- 4.6. While law can be complicated, giving people access to the most basic of legal services and insight from 'robot lawyers' need not be if the first step is to use AI tools and chatbots wherever possible.
- 4.7. Do people really want to be talking to 'robot lawyers' instead of human lawyers? According to a recent study of UK consumers commissioned by Ubisend, when asked "Why would you consider asking a chatbot before a human?" nearly 70% said they would prefer to receive an instantaneous answer. When asked "What's most important when communicating with a company, 68% say it is 'reaching the desired outcome', closely followed by 'ease of experience' (48%) and 'speed' (44%). Being able to contact a brand at the time most convenient to them is also important to 39% of UK consumers.
- 4.8. Cranfield presents the business perspective. Outsourcing decision-making to machines risks unintended discriminatory outcomes. As technology firms increasingly exploit the possibilities of machine learning, they could become vulnerable to legal pitfalls both now and increasingly so in the future. Tech businesses may need to take steps to protect themselves from potentially thorny legal issues. This may be a growing area of business for legal firms.
- 4.9. The impact on workforce mental health and wellbeing is increasingly recognised as a significant issue as AI starts to change the nature of work and jobs. How will today's leaders and managers respond to these psychosocial and health effects of work reorganisation? Equally, how will legal firms providing advice to providers of mental health services in the statutory and third sector anticipate changing needs? Larger companies are already looking ahead, as is government, in planning for the likely social and health impact and the consequences for legal firms.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?
 - 5.1. In order to address the misunderstanding/lack of understanding of the regulatory framework by users of AI tools and systems with regard to the existing provision and future provision, Lightfoot suggests that perhaps there should be a 'professional' standard for AI legal service providers. For example, an AI legal service provider should:
 - be explicit in communicating about its technology to the user so that the user understands what kind of AI it is, how it works, and whether, or at what stage in the process a human is involved;
 - consider appropriate levels of transparency in how they use AI to interact with customers;
 - Provide clarity on how AI benefits the customer experience;

- inform customers about the provisions in place to safeguard their data.

The above would help in relation to the positive impact by AI on society with regard to AI legal service providers having to comply with an 'AI legal regulation' that the user can understand and draw comfort from (and thereby help the AI legal service provider 'sell into' / provide free services to businesses and consumers) whilst at the same time educate the buyers / users of such AI tools that this regulation is in place to safeguard everyone's interest (and thereby help break down the perception of AI and use of AI barriers that exist currently, which may encourage potential buyers / users of AI legal AI tools to begin to use these positive alternative solutions).

- 5.2. Education has a role to play. Cranfield outlines the approach of Business and Law Schools in London: Westminster Law School has set up a hub in Law and AI. In terms of robotics, much of the research is based in informatics, e.g. [CORE at Kings](#). Several universities offer courses in Artificial Intelligence, Automation Control, Computational Intelligence, Cognitive Robotics, and Intelligent Systems, some of which offer work placements with blue-chip organisations.
- 5.3. This summer, London has seen open days, public events and exhibitions, showcasing the work of AI and robotics developers and researchers from universities and the private sector, e.g. at King's College London, Imperial College, the Science Museum, and the Royal Society. These have been incredibly popular with the general public, especially families. Westminster Law School has hosted a film series and a series of panel discussions.

Ethics

6. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?
 - 6.1. The human experience of obtaining legal advice and interacting with the justice system should not be forgotten when designing new processes or introducing AI into existing processes. Legal services and the justice system do not simply deliver a transactional outcome where words are processed in order to resolve an issue. They involve complex human relationships, both between individuals, between individuals and corporations or between individuals and the state. Concepts such as trust, empathy, fairness, justice and other emotions will continue to form a key part of people's expectations when obtaining legal services or interacting with the justice system.
 - 6.2. These human considerations must not be undermined by AI, and ought to be enhanced. The risk is that developments purely to obtain cost-savings or efficiencies may undermine access to justice, ethics and

trust in the law (see recent Supreme Court decision on Employment Tribunal Fees with similar considerations).

7. In what situations is a relative lack of transparency in artificial intelligence systems (so- called 'black boxing') acceptable? When should it not be permissible?
 - 7.1. Transparency and data are key issues, particularly in the legal sector where decisions or advice are often based on sensitive personal information. Access to legal services and to the justice system is not simply a commodity to be consumed, but is sacrosanct and determines the parameters of the individual's relation to the state. Similarly to the 'data protection by design' principle in the General Data Protection Regulation, AI for use in law should be designed with trust, transparency and accountability as foremost concerns. In the legal sector, higher standards of accountability and transparency, as reflected in the professional codes of conduct, must be expected.

The role of the Government

8. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?
 - 8.1. Processing of personal data is already regulated in the UK by the Information Commissioner's Office (ICO) who published an excellent paper on artificial intelligence, machine learning and big data in March 2017. In this paper, the ICO considers many issues posed by AI in relation to data protection legislation (including the upcoming General Data Protection Regulation). The requirements for transparency, accountability and ultimately fairness are challenged by AI, but different solutions will be needed in different sectors.
 - 8.2. Many of the legal services that could be provided using AI will involve some processing of personal data and therefore be regulated by the ICO. The Government needs to review and consider its regulatory approach in the light of new technologies, and how the ICO can work with regulators (e.g. the SRA and the Bar Standards Board) and other public bodies in other sectors. Recent experience with the financial services sector has shown that regulators must have the technical and commercial insight to keep-up with industry developments and significant and meaningful investigatory and enforcement powers, but without having a revolving door between the industry and the regulator. This may be very difficult to implement in practice, particularly when AI is implemented across different industries (some of which have existing regulators), and needs to be considered carefully.

6 September 2017

Google – Written evidence (AIC0225)

Google UK written submission to the House of Lords Select Committee on Artificial Intelligence inquiry into the economic, ethical and social implications of advances in artificial intelligence

1.1 Google welcomes the Lords Select Committee on Artificial Intelligence (AI) inquiry into the economic, ethical and social implications of advances in AI, and appreciates the opportunity to provide input based on our experience.

1.2 Google’s mission is to organise the world’s information and make it universally accessible and useful. In pursuing this mission, we have always been excited by the promise of AI and the real life benefits it can impart to society. Driven by significant advances in research and computing power, this technology, particularly in the application of machine learning (ML), is increasingly becoming a key feature of our products.

1.3 More generally, we believe ML is a flexible, powerful tool that could provide crucial help in advancing science, improving access to medical care, and tackling some of the biggest global challenges. However, like any technology, there is nothing predestined about the impact of AI and ML. People will make choices about how and where to implement these technologies, and in doing so shape their influence on society. Now is a timely moment to be having this discussion, in order to mitigate risks and maximise benefits of this technology for everyone.

2. The pace of technological change

- What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?
- Is the current level of excitement which surrounds artificial intelligence warranted?

2.1 AI is not a speculative science fiction technology but a practical software engineering tool already being used to help millions of people around the world every day. Machine learning, a field within AI that specifically studies algorithms that learn from data, is already benefiting many of Google’s products.

2.2 Recent breakthroughs in AI have been decades in the making. Many contemporary developments draw heavily on insights from research carried out in the 1980s and 1990s, which are only now becoming practical because of the availability of computational power, richer sources of information about the world, and a growing community of talent. However, what has clearly accelerated in recent years is the translation of AI research to real world problems. We appear to have reached a tipping point at which the pace of

change in the lab is being matched by its practical application.

AI's use in our everyday lives

2.3 In our products, ML advances have boosted efforts ranging from user protection -- with improved spam and malware filters -- to enhanced accessibility, through stronger voice recognition. For example, using a method⁵⁶⁸ that learns language from patterns in bilingual text, Google Translate⁵⁶⁹ translates more than 100 billion words a day in 103 languages. AI has been key in helping us achieve significant breakthroughs with speech recognition⁵⁷⁰, reaching near-human levels of accuracy. With Google Photos⁵⁷¹, you can search for anything from "hugs" to "border collies" because the system uses our latest image recognition system to automatically categorise objects and concepts in images. We also recently announced that ML helped optimise system settings to cut energy consumed cooling our data centres by up to 40%⁵⁷².

2.4 Beyond enhancing existing products, these technologies will drive efficiencies and may dramatically improve society's ability to tackle some of our most pressing global challenges in health, environment, transportation, and beyond. For example, last year Google showed how ML can make the diagnosis of diabetic retinopathy⁵⁷³- one of the fastest growing causes of blindness worldwide - more broadly accessible.

AI's use in the future

2.5 One of the key reasons it's hard to make progress on important social challenges in medicine, energy, and science is that even the smartest experts struggle to fully understand the relationships between cause and effect in these systems. Machine learning technology has the ability to parse the complexity of interacting factors and volume of information, and in turn allow us -- and many others -- to design more effective interventions.

⁵⁶⁸ Inside Google Translate, [www.youtube.com/watch?v= GdSC1Z1Kzs](http://www.youtube.com/watch?v=GdSC1Z1Kzs)

⁵⁶⁹ Barak Turovsky, Ten Years of Google Translate, April 2016, www.blog.google/products/translate/ten-years-of-google-translate/

⁵⁷⁰ Barb Darrow, Fortune: Google Says Its Speech Recognition Leads the Pack, May 2017, for.tune.com/2017/05/18/google-speech-recognition/

⁵⁷¹ Francois de Halleux, Smarter photo albums, without the work, March 2017, blog.google/products/photos/smarter-photo-albums-without-work/

⁵⁷² Richard Evans & Jim Gao, DeepMind AI reduces energy used for cooling Google data centres by 40%, July 2016,

blog.google/topics/environment/deepmind-ai-reduces-energy-used-for/

⁵⁷³ Lily Peng, Detecting diabetic eye disease with machine learning, November 2016,

blog.google/topics/machine-learning/detecting-diabetic-eye-disease-machine-learning/

2.6 Researchers are exploring the use of ML on topics ranging from weather prediction⁵⁷⁴ and genetic diseases⁵⁷⁵ to conservation⁵⁷⁶ and economic forecasting.⁵⁷⁷ Innovators are just starting to build out practical use cases, and we are excited to see the results.

3. Impact on society

- How can the general public best be prepared for more widespread use of artificial intelligence?
- Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Focus on employment

3.1 Throughout history new technologies have always played a role in shaping the nature of employment, and we should expect that increased use of AI and ML will be no different. In many sectors, ML will augment and enhance the work that people do, enabling them to be more effective in the same roles - boosting productivity. As with all technological innovation, we should expect that new areas of economic activity and employment will be made possible, and some types of work and some skills will decrease in relevance. More generally, in an era when many countries are facing falling productivity growth, an aging workforce, and increased global competition, AI could be an important means of addressing critical challenges in economies around the world.

3.2 One of the most important steps we must take so that everyone can benefit from the promise of AI is to ensure that current and future workforces are sufficiently skilled and well-versed in digital skills and technologies, particularly STEM subjects. The UK government has been proactive and vocal in support for digital education, such as introducing computer science into the curriculum from 2014, but it is important not to be complacent about the step change that is needed. Any advance in technology must be met by an improved curriculum and teachers' ability to deliver it.

3.3 To ensure we have teachers that are fully trained to effectively deliver the new coding curriculum Google has partnered with Teach First to help support and train the next generation computing teachers to specifically address the acute teacher shortage in this subject area, but the scale of the challenges means that

⁵⁷⁴ McGovern et al, Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning, April 2013,

<https://www.ncbi.nlm.nih.gov/pubmed/26549932>

⁵⁷⁵ Anshul Kundaje, Current Research and Scholarly Interests, med.stanford.edu/profiles/anshul-kundaje

⁵⁷⁶ Tanya Y. Berger-Wolf, Computational Population Biology at UIC, compbio.cs.uic.edu/

⁵⁷⁷ National Science Foundation, Machine learning and the wisdom of the crowd, April 2016,

https://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=138021&org=N SF

large scale investment by government is needed.

3.4 We urgently need to address the digital skills gap by focussing on education, teacher supply, adult skills and digital inclusion, as well as investing in digital and creative skills wherever possible.

3.5 It is also important that the UK is able to harness the talents of the widest pool available, which means putting real effort into encouraging more women into technology, focussing on adult digital literacy as well as youth education, and enabling the next generation of entrepreneurs no matter their socioeconomic background. It is clear that the technology industry faces a problem of gender disparity that can be traced back to the relatively small numbers of girls who take up STEM subjects at school and university. We welcome the work of Martha Lane Fox and Doteveryone in enabling technologies that advantage all British citizens.

3.7 Beyond formal education, there is also the challenge of ensuring that the existing workforce has opportunities for continued education and reskilling to leverage the latest improvements in technology. Providing access to and motivation for lifelong learning is one of the most crucial things that can be done to prepare for future developments.

Ensuring the benefits of AI are shared by all

3.8 As with any technology, it is important to maximise the positives and minimise any possible harms. Although in everyday terms the benefits are already being felt by millions of users through improved products and services, it is vital that we are mindful of and take action to minimise the risk that AI and ML entrenches existing inequalities in society.

3.9 At Google we believe a key part of ensuring AI is useful to all of us is to build systems with people in mind at the start of the process. We recently People + AI Research initiative (PAIR)⁵⁷⁸ which brings together researchers across Google to focus on the "human side" of AI: the relationship between users and technology, the new applications it enables, and how to make it inclusive.

3.10 PAIR's research is divided into three main areas, based on different user needs:

- **Everyday users:** How might we ensure AI and ML is inclusive, so everyone can benefit from breakthroughs in AI? Can design thinking open up entirely new AI applications? Can we democratise the technology behind AI?
- **Engineers and researchers:** AI is built by people. How might we make it easier for engineers to build and understand ML systems? What

⁵⁷⁸ People+AI Research Initiative, <https://ai.google/pair>

educational materials and practical tools do they need?

- **Domain experts:** How can AI aid and augment professionals in their work? How might we support doctors, technicians, designers, farmers, and musicians as they increasingly use AI?

4. Public perception

- Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

4.1 Public perception, trust and engagement with AI must be improved if the industry is to continue to thrive. This is particularly important for the UK if it wishes to maintain its position as an academic and industrial world-leader in AI and ML. The rate and scale at which AI and machine learning technology impacts society will be influenced by a multitude of factors - technological advancement being only one. Google is committed to playing our part in boosting public engagement with AI, but, for the greatest chance of success, this needs to be a multi-stakeholder effort. This is why Google was a founding member of the Partnership of AI, bringing together leaders in the tech sector, civil society and academia.

4.2 Clearly, data relevant to the problem you are trying to solve is an essential ingredient for training ML models. At Google we are working to raise the bar on data privacy and transparency, to ensure public confidence in the use of data for such purposes. For example, using My Account, a website that lets you intuitively control what you share, users can see their information like web and app activity, their location information, or their YouTube watch history. They can control what data gets associated with their Google account, pause the collection of specific types of data, or delete a specific entry, day, or even their entire history, including from Google Home.

4.3 New technologies necessarily challenge us to think of new ways to protect our users' privacy in practice. These are, in essence, engineering and design challenges, and we are at the forefront of exploring them. For example, for the subset of machine learning models trained based on data from user interactions with mobile devices, we're researching Federated Learning in which model training takes place on the device. Looking further ahead, we're also exploring generative machine learning, in which 'synthetic' data is used for models to learn from — i.e.: data that realistically reflects the world, but is not real data and thus would not present privacy concerns.

4.4 We are also working at improving transparency in AI decisions. For example, trade-offs in ML and AI can be highly complex and so, at PAIR, we are working on ways to explain fairness, with visualisations and interactive explanations to help people understand these critical issues. For example, we recently produced

[a video](#)⁵⁷⁹ highlighting the challenges of human bias in ML.

5. Industry

- What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?
- How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

5.1 All sectors of the economy stand to benefit from the development and use of artificial intelligence. As part of our efforts to make sure these benefits are felt across different sectors and by businesses large and small we offer Google Cloud AI.⁵⁸⁰ Cloud AI provides modern machine learning services, with pre-trained models and a service to generate to businesses' own tailored models. Cloud AI uses the same cloud machine learning as major Google applications such as Photos, Translate and voice search, bringing unmatched scale and speed to business applications.

5.2 AI has flourished in part because of a set of common norms that encourage research results to be published and shared openly. It is important to preserve these community principles towards openness that have proven important to past work in the space.

5.3 In this vein, Google has been an active and open contributor to the [research community](#).⁵⁸¹ We have [published results](#)⁵⁸² and actively participated in conferences on a variety of topics including large scale deep learning, computer vision, sequence to sequence modeling, and visualisation of the internal processes of neural networks.

5.4 We're also releasing open source tools for researchers and other experts to use. For example, in November 2015 we open sourced TensorFlow⁵⁸³ - Google's internal ML toolkit to allow anyone to experiment in the space and advance the state of the art. More recently, we released [FACETS](#)⁵⁸⁴, a set of data visualisation tools to aid in understanding and analysing the datasets used to train ML models.

5.5 Many advances in machine learning have been made with the use of publicly

⁵⁷⁹ Machine Learning and Human Bias, [youtube.com/watch?v=59bMh59JQDo](https://www.youtube.com/watch?v=59bMh59JQDo)

⁵⁸⁰ Google Cloud Platform, Cloud AI, cloud.google.com/products/machine-learning/

⁵⁸¹ Research at Google, Publications, research.google.com/pubs/papers.html

⁵⁸² *Ibid.*

⁵⁸³ Tensor Flow, [tensorflow.org/](https://www.tensorflow.org/)

⁵⁸⁴ Facets, <https://pair-code.github.io/facets/>

available datasets. For example, datasets such as ImageNet⁵⁸⁵, COCO⁵⁸⁶ and YFCC100M⁵⁸⁷ have been crucial to progress in computer vision. Google is committed to contributing to this ecosystem, through our investment in gathering and preparing datasets for open source release. A fuller list is [here](#)⁵⁸⁸ but in 2017 alone we have released large datasets supporting researchers across a wide range of fields, including [speech commands](#)⁵⁸⁹, [photos23](#)⁵⁹⁰ and [video24](#)⁵⁹¹, [online discussion](#)⁵⁹², [audio effects](#)⁵⁹³, and [crowdsourced drawings](#).⁵⁹⁴

5.6 Many companies have proven that you can quickly outperform competitors that have much more data than you, just by having a better product or service. In fact, it is the expertise and ability to interrogate data to derive value from it that matters much more than raw ingredients. The competitive landscape will center around which companies can develop the best tools to derive useful insights from data, rather than simply which ones own the largest quantity of data.

5.7 There may be a role for government to consider how it can expand the use of data analytics and other tools by support the development of these skills as part of its digital strategy.

Cyber-security

5.8 Ensuring the highest standards of data security is key to ensuring it contributes to the public good and a well-functioning economy. Managing data

⁵⁸⁵ Image Net, <http://image-net.org/>

⁵⁸⁶ Common Objects in Context (COCO), <http://cocodataset.org/>

⁵⁸⁷ Yahoo Research, Yahoo Flickr Creative Commons, <webscope.sandbox.yahoo.com/catalog.php?datatype=i&did=67>

⁵⁸⁸ Google Research Blog, Datasets, <research.googleblog.com/search/label/datasets>

⁵⁸⁹ Pete Warden, Launching the Speech Commands Dataset, August 2017, <https://research.googleblog.com/search/label/datasets>

⁵⁹⁰ Vittorio Ferrari, An Update to Open Images - Now with Bounding-Boxes, July 2017, <research.googleblog.com/2017/07/an-update-to-open-images-now-with.html>

⁵⁹¹ Paul Natsev, An updated YouTube-8M, a video understanding challenge, and a CVPR workshop. Oh my!, February 2017, <research.googleblog.com/2017/02/an-updated-youtube-8m-video.html>

⁵⁹² Praveen Paritosh & Ka Wong, [Coarse Discourse: A Dataset for Understanding Online Discussions, May 2017,](#) [research.googleblog.com/2017/05/coarse-discourse-data set-for.html](research.googleblog.com/2017/05/coarse-discourse-data-set-for.html)

⁵⁹³ Dan Ellis, [Announcing AudioSet: A Dataset for Audio Event Research, March 2017,](#) [research.googleblog.com/2017/03/announcing-audioset-data set-for-audio.html](research.googleblog.com/2017/03/announcing-audioset-data-set-for-audio.html)

⁵⁹⁴ Reena Jana & Josh Lovejoy, [Exploring and Visualizing an Open Global Dataset, August 2017,](#) <research.googleblog.com/2017/08/exploring-and-visualizing-open-global.html>

securely is critical to being able to continue to improve the apps and services we all rely on with AI and ML. With UK citizens beginning to see the benefits of big data, data protection questions remain key to building and maintaining public trust, especially with a number of public services and organisations using different security protocols to share data.

5.9 As secure and protected ways of providing data continue to evolve, government should play a significant role in supporting academic research into world-leading data security practices that would be widely adopted in the UK. Secure data will be one of the key foundations upon which success in AI research and innovation is built, as it will maintain public trust.

6. Ethics

- What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?
- In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

6.1 As with all scientific research, ethical oversight is important. As previously mentioned, there are key ethical and safety concerns around the security of data. These concerns require attention both by researchers and by government. The recent rapid advance of AI and ML has also raised concerns surrounding the safety of implementing these systems in a variety of different contexts.

6.2 Google is committed to advancing the use of AI and ML in an open and ethical way. No system is perfect, and errors will emerge. These errors are diverse and we should not expect that they will be resolved through a simplistic "one size fits all" solution. However, over time we expect advances in our technical capabilities will expand our ability to meet these challenges.

Ethics in theory

6.3 To that end, we believe that solutions to these problems can and should be grounded in rigorous engineering research to provide the creators of these systems with approaches and tools they can use to tackle these problems. "[Concrete Problems in AI Safety](#)"⁵⁹⁵, a recent paper from our researchers and others, takes this approach laying out 5 key questions around safety for researchers to tackle. Google is committed to the responsible development of AI and we are among the founding partners of the [Partnership on AI](#)⁵⁹⁶, an independent non-profit organisation being created to study and formulate best practices on AI technologies to ensure it brings positive benefits to society.

⁵⁹⁵ Chris Olah, [Bringing Precision to the AI Safety Discussion](https://research.googleblog.com/2016/06/bringing-precision-to-ai-safety.html), June 2016, research.googleblog.com/2016/06/bringing-precision-to-ai-safety.html

⁵⁹⁶ Partnership on AI, <https://www.partnershiponai.org/>

6.4 We also believe that graduate degrees within computer science should incorporate mandatory ethics courses along the same lines as the ethics training required for medical and legal qualifications, including training in the ethics of data science and algorithmic fairness.

Ethics in practice

6.5 Potential harms are not just matters of research. As a recent ProPublica⁵⁹⁷ investigation into ML used by the judicial system in the US illustrated, partial or biased data can produce discriminatory results as ML algorithms draw incorrect inferences from the examples they are trained on. Equally, by focusing on more objective criteria, ML might help reduce or avoid discrimination.

6.6 In his article, [Equality of opportunity in supervised learning](#)⁵⁹⁸, Google researcher Moritz Hardt⁵⁹⁹ looked at the short term questions of bias and discrimination. In the article Hardt points to the need for improved tools for diagnosing these failures, as well as the need to avoid [data gaps](#)⁶⁰⁰ where the [dearth of good data](#)⁶⁰¹ can make the use of ML problematic.

6.7 Ultimately, as with any advanced technology, the impact of AI will reflect the values of those who build it. Mathematical and technical constraints force tradeoffs that practitioners and policymakers will need to resolve between accuracy and interpretability or fairness. AI is a tool that we humans will design, control and direct. It is up to us all to direct that tool towards the common good.

Focus on transparency

6.8 Overcoming the trade-off between interpretability and performance for complex ML models is among the most researched areas in the field today. Advancing this is a priority for Google, not only because it is key to boosting trust in the results of such models, but also because it is likely to yield insights that lead to further improvements.

6.9 Examples of our work in this field include:

⁵⁹⁷ Angwin et al, Machine Bias, May 2016, propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

⁵⁹⁸ Moritz Hardt, Equality of Opportunity in Machine Learning, October 2016, research.googleblog.com/2016/10/equality-of-opportunity-in-machine.html

⁵⁹⁹ Moritz Hardt, <http://www.moritzhardt.com/>

⁶⁰⁰ Daniel Castro, The Rise of Data Poverty In America, September 2014, <http://www2.datainnovation.org/2014-data-poverty.pdf>

⁶⁰¹ Melinda Gates, To Close the Gender Gap, We Have to Close the Data Gap, May 2016, <https://medium.com/@melindagates/to-close-the-gender-gap-we-have-to-close-the-data-gap-e6a36a242657>

- **Distill:** In partnership with Open AI, DeepMind, and others we have established Distill, an independent organisation to support a new open science journal and ecosystem supporting human understanding and clarity in machine learning.
- **Deep Dream:** In its earliest incarnation this project was aimed at visualising what different layers within a neural net were learning during training, to make it easier to spot where mistakes in classification arose.
- **Glassbox** is a machine learning framework optimised for interpretability. It involves creating mathematical models to smooth out the influence of outliers in a data set, thus helping to make results more predictable and decipherable.
- **FACETS** tool aims to help engineers have a clearer view of the data they use to train AI systems so as to better understand and debug what they're building

6.10 We're optimistic that these efforts will provide us with clearer explanations over time, even if there are limits to what is possible now. While some systems will take more effort than others to 'cut open', with enough resource it's already possible to learn something about how even the most complex models work.

7. The role of Government

- What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

7.1 In order for the UK to remain as a world leader in AI and ML the Government must continue to support the growth of the Digital Economy. Other countries are developing their AI capabilities at a rapid pace, and the UK must remain competitive.

7.2 The Government has a vital role to play in creating an environment in which private research can thrive, for example by maintaining a pro-innovation legal regulatory regime which provides clarity and flexibility, along with appropriately nuanced safeguards. To this end, we welcome the upcoming publication of Wendy Hall's review into AI in the UK.

7.3 We believe that government has an important contribution to make in promoting and supporting the development of AI technologies. In particular, we believe the Government should:

- Support research: The Government has traditionally played an important role in supporting long term fundamental research. The government has and should continue to play that role with AI, supporting research into the novel application of these technologies in meeting social challenges and addressing potential limits and shortfalls. The Government should also convene a panel of academic and industry experts to determine research

funding priorities and directions with an emphasis on transparency and accountability, and feed these through to research councils and other funding bodies.

- Fund AI masters and PhDs: The Government should consider funding for AI programmes at British universities, to encourage more research in the field and nurture the next generation who will help preserve the UK's preeminent position. This funding could also include direct support for modules within programmes that train researchers in the ethics of data science, to ensure that the pursuit of beneficial outcomes is embedded in the science of AI at every level.
- Preserve Open Data norms: Open research norms and practices have encouraged research results to be published and shared openly and enabled AI to flourish in the UK. We urge the government to help preserve and encourage these community principles of openness, for instance by promoting the release of complete, high quality and robust public datasets for use by the wider research community.
- Support clusters: The UK can only have a world-class AI sector if it ensures a world-class tech cluster is already in existence. Google DeepMind, a British AI company acquired by Google in 2014 is based in The Knowledge Quarter⁶⁰², a world-class knowledge cluster in the heart of London that contains some of the world's leading scientific institutions including the British Library⁶⁰³, Central St Martins⁶⁰⁴, The Francis Crick Institute⁶⁰⁵ and The Alan Turing Institute⁶⁰⁶, allowing unrivalled opportunities for collaboration and learning. The government should consider how it can build on this success and increasing the number of science-led organisations in the King's Cross area so that a scientific cluster is allowed to flourish.
- Promote Education and Diversity: The government has encouraged public support for increasing access to and diversity in STEM education and careers. We should continue this effort to ensure that more students from more backgrounds have access to computer science education. We strongly support the Computer Science for All Initiative, and would encourage similar projects. Google also believes that having people from a variety of perspectives, backgrounds, and experiences working on and developing the technology will help to identify potential issues.

⁶⁰² The Knowledge Quarter, <http://www.knowledgequarter.london/>

⁶⁰³ The British Library, <https://www.bl.uk/>

⁶⁰⁴ Central St Martins, <http://www.ar ts.ac.uk/csm/>

⁶⁰⁵ The Francis Crick Institute, <https://www.crick.ac.uk/>

⁶⁰⁶ The Alan Turing Institute, <https://www.turing.ac.uk/>

- Ensure Flexibility: Despite many recent breakthroughs, AI and its applications are still nascent. We recognise the need to build public trust and understanding, but as these opportunities are still emerging, we encourage a cautious and nuanced regulatory approach that will allow innovative uses to flourish and reach their full potential. It's vital not to enshrine a set of technical designs or methods that might be obsoleted by new methods that may be better at preserving privacy, safety, and interoperability, among other values. In consumer protection areas like privacy, it will be important for existing agencies to maintain a harmonised approach as they assess whether new rules are needed and, if so, how they should be integrated with existing approaches developed over time. We also believe consensus driven best practices and self-regulatory bodies will play an important role in ensuring the flexibility necessary to drive innovation while simultaneously developing nuanced and appropriate safeguards.
- Convene Talent to Meet Social Challenges: Machine learning has proven to be an effective tool for making progress on complex problems at significant scale. The government faces these types of challenges in fields like energy, transportation, environment, urban planning, and public health. We believe that it can convene task forces to explore the use of AI in these fields and to improve the work of Government.

8. Learning from others

- What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

8.1 It's important to note that not everything that AI/ML makes possible requires a new set of rules — in the vast majority of instances, existing frameworks will be sufficient for ensuring the protection of key values. In the rare cases that new rules may be needed to address safety, operations, or other product infrastructure areas, we support the approach of having expert agencies take the lead on regulation of specific uses in their areas — such as illustrated by the approach to self driving cars.

8.2 More generally, when it comes to the issue of responsible development of AI, we believe that an open, participatory, interdisciplinary process with representatives from civil society and private industry is needed to create a credible set of self-regulatory principles. In this regard, we support the approach taken by the Japanese government to date in the development of its draft principles.

17 October 2017

Government of Canada – Written evidence (AIC0222)

Introduction

The House of Lords' Select Committee on Artificial Intelligence was appointed to consider the economic, ethical and social implications of artificial intelligence. As part of a public call for evidence, the Committee invited the Government of Canada to submit evidence to its inquiry. In particular, the Committee invited Canada's views on the following questions:

- Q1: The role of government and regulation when making policy on artificial intelligence; and
- Q2: Canada's contribution to fostering international efforts or initiatives on artificial intelligence.

This document presents Canada's current position with regards to these questions. The document was prepared by Innovation, Science and Economic Development Canada with contributions from other departments and agencies from the Government of Canada.

Q1: The role of government and regulation when making policy on artificial intelligence

Response to question 1: the Government of Canada has not yet conducted a structured and comprehensive pan-government reflection on what its role, or the role of governments in general, should be with regards to policy- and regulation-making around artificial intelligence.

However, a number of Government of Canada departments and organizations are already conducting AI-related activities within their spheres of responsibility. These activities include supporting fundamental research and training efforts in AI, leveraging AI to drive growth across Canada's industries as part of the government's economic development policy and undertaking continuing policy foresight activities. In addition, several federal organizations are experimenting with applications of AI technologies to assess the potential benefits and risks of using these technologies in their operations. This work mostly focuses on the technical aspects of AI technologies but discussions have been initiated about potential policy or regulatory frameworks that could support the application of the technologies. These discussions are still at a very early stage. More detail about some of the activities underway is provided below.

Innovation, Science and Economic Development Canada

Innovation, Science and Economic Development Canada's mission is to foster a growing, competitive, knowledge-based Canadian economy. The Department works with Canadians throughout the economy, and in all parts of the country, to improve conditions for investment, improve Canada's innovation performance, increase Canada's share of global trade, and build an efficient and competitive marketplace. Innovation, Science and Economic Development Canada's mandate is organized into three interdependent and mutually reinforcing strategic outcomes: advancing the marketplace; fostering the knowledge-based economy; and supporting business.

Thanks to a number of early and substantial investments by the Canadian federal government over the past 10 years, and to the pioneering work done by outstanding Canadian researchers, Canada has developed recognized capacity and expertise in the area of artificial intelligence. Canada's leadership is particularly strong in promising AI subfields such as deep learning and reinforcement learning, two approaches to artificial intelligence that are of high – and rising – interest to industry, government and others for their potential to lead to practical applications. The Innovation and Skills Plan, introduced in Budget 2017, is Canada's agenda to become a world-leading centre for innovation, help create better, well-paying jobs, and help strengthen and grow the middle class. The Innovation and Skills Plan identified artificial intelligence as a key platform technology that will drive growth across Canada's industries and as an area where the country has the potential to be a global leader.

An initiative announced in the Innovation and Skills Plan, the **Pan-Canadian Artificial Intelligence Strategy**, which received \$125 million investment in Budget 2017, aims to promote collaboration between Canada's main centres of expertise in Montreal, Toronto-Waterloo, and Edmonton. This investment is intended to position Canada as a world-leading destination for companies seeking to innovate through the application of artificial intelligence technologies by helping to retain and attract top academic talent in the field of AI and increase the number of post-graduate trainees and researchers in this area.

In addition, the Strategy highlights that artificial intelligence is expected to have profound implications for the economy, government and society. As such, the Strategy includes an investment to support the engagement of eminent researchers in policy-relevant working groups that examine the breadth of implications of AI, and the publication of their findings to inform the public and policy-makers. It is expected that this investment will help Canada develop global thought leadership on the economic, ethical, policy and legal implications of advances in AI.

Policy Horizons Canada

Policy Horizons Canada (Horizons) is a foresight organization within the Government of Canada whose mandate is to help anticipate emerging policy challenges and opportunities, explore new knowledge and ideas, and experiment with methods and technologies to support resilient policy development. Horizons has identified AI as a change driver that will disrupt the economy and society over the next 15 years. In particular, Horizons has explored AI in the following areas:

- plausible futures of AI over the next 15 years;
- analysis of deep learning, what it can and can't do today;
- interaction of AI, robots, task unbundling, and telework with respect to the changing nature of work and plausible future economies in 2030; and
- impact of AI on citizens' expectations with respect to service delivery and, as a result, the way government may be expected to operate.

[For AI publications, please refer to (<http://www.horizons.gc.ca/eng/content/publications>)]

In addition, Horizons, in collaboration with the Treasury Board of Canada Secretariat, has initiated with the policy community an interdepartmental dialogue on AI to discuss policy and regulatory challenges and opportunities that could emerge.

National Research Council of Canada

The National Research Council (NRC) is the Government of Canada's premier research organization supporting industrial innovation, the advancement of knowledge and technology development, and fulfilling government mandates.

Building on its deep learning and machine learning expertise, the NRC is developing a program in AI to increase the availability of AI technologies and expertise to Canadian companies and government, with the goal of promoting and maintaining Canada's leadership in the field and continuing to grow the Canadian industry. The program will work collaboratively to connect the Canadian AI eco-system in support of:

- Advancement of knowledge: developing next generation AI in collaboration with academia.
- Development of applications: translating knowledge into applications in collaboration with the Canadian industry and government departments.
- Use of AI for the public good: translating knowledge and market applications for regulators in support of regulatory needs for the secure, reliable, safe and ethical use of AI.

Department of Justice Canada

The Department of Justice Canada has the mandate to support the dual roles of the Minister of Justice and the Attorney General of Canada. The Department also works to ensure the federal government is supported by high-quality legal services, and the justice system is fair, relevant, accessible, and reflective of Canadian values.

The practice of law is changing. New technologies and new ways of working are transforming the legal profession and how legal services are delivered. AI technologies have the potential to significantly impact the Department of Justice's activities. Web-based legal information systems, document review, predictive analytics, legal research, online dispute resolution tools and general access to justice are examples of areas that could see fast and significant advances thanks to AI technologies. As highlighted by the Select Committee, advances in AI also raise a number of questions at the intersection of technology, law and ethics. For this reason, Justice Canada has recently created a Task Force on AI with the mandate to:

- Assist Justice Canada to understand the current state of AI developed for the legal profession; its existing and potential applications; and the challenges and opportunities that it may present;
- Assess the potential benefit, and potential implementation challenges, of existing commercially-available AI solutions for the activities of the Department of Justice and for other legal activities in the federal government. As part of the assessment, reviews of existing products and small-scale pilot experiments will be conducted;
- Identify anticipated legal and ethical issues related to the use of AI and initiate reflections on the development of a legal framework for AI;
- Analyze the opportunity for the Department of Justice to develop its own in-house technical capacity for the development of AI solutions and
- Review the current legal framework for AI in Canada, identify questions that are likely to require legal developments in the near future, and identify the expected needs for legal expertise in the Government of Canada.

Treasury Board of Canada Secretariat

The Treasury Board of Canada Secretariat provides advice and makes recommendations to the Treasury Board committee of ministers on how the government spends money on programs and services, how it regulates and how it is managed. The Secretariat helps ensure tax dollars are spent wisely and effectively for Canadians.

Increasingly, federal government institutions are looking to AI to improve the delivery of services and the design of public policy. As a central agency of the Government of Canada, the Treasury Board of Canada Secretariat is working

with all interested institutions to ensure that this is done in a coherent, responsible, and ethical fashion. To this effect, the Secretariat is in the process of developing non-binding ethical guidance through a Government of Canada white paper on the responsible use of AI for policy and services, which is expected to be completed by December 2017.

In tandem, the Secretariat will be working with federal institutions and key partners such as the National Research Council (NRC) to test AI- and algorithmic-based approaches to service design and delivery. This work will be done with an aim to provide tangible guidance on implementation to institutions. Ethical guidance in development will be informed by these tests, and *vice versa*.

Q2: Canada's contribution to fostering international efforts or initiatives on artificial intelligence

Response to question 2: Canada has been an active participant in international forums where discussions have touched on artificial intelligence. These include the G7 Industry and ICT Ministers forum, the G7 Leaders forum and the OECD. So far, Canada has advocated the importance of taking a human-centric approach to the development and deployment of AI and has highlighted the need for further discussion and collaboration between states in understanding the opportunities and challenges brought by AI and identifying areas where actions could be considered. Canada is also taking steps to better understand the potential challenges and opportunities that AI poses for Canadian foreign policy objectives related to human rights, inclusion and democracy.

Innovation, Science and Economic Development Canada

Canada is an active participant to the G7 Industrial and ICT Ministers forum and the Minister of Innovation, Science and Economic Development is the Canadian representative at the forum. The forum has taken an interest in AI as a strategically important emerging technology that has great potential to boost economic productivity and improve social well-being. To date, Canada has advocated the importance of taking a human-centric approach to the development and deployment of AI. Specific considerations raised by Canada include:

- As a leader in AI, Canada shares experiences in supporting technology development in an open, transparent and inclusive manner.
- Public trust and acceptance are necessary for the widespread deployment of emerging technologies. Canada sees government as having a critical role to play in facilitating innovation and growth as well as ensuring that workers are prepared for the jobs of the future and get the training they need to succeed in the new economy.

- International engagement is important to Canada. AI is a relatively new technology and research and dialogue between countries will help deepen our understanding of the multifaceted opportunities and challenges it brings and develop a perspective that is neutral to any advancement in AI. International engagement is also necessary to explore the opportunity of developing multi-stakeholder approaches to policy and regulatory issues that include technical and societal considerations posed by AI.

Global Affairs Canada

Global Affairs Canada's (GAC) Office of Human Rights, Freedoms and Inclusion (OHRFI) is seeking to support the Government of Canada's commitment to being a global innovator in AI development, while ensuring Canada continues to be a responsible stakeholder that is committed to meeting international human rights obligations. In this context, the OHRFI-hosted Digital Inclusion Lab is exploring how Canada can support the development of AI in a way that enhances and promotes international human rights. Specifically, the Lab is engaging multilateral fora related to human rights and digital freedom to encourage dialogue around the role of governments, private sector, and civil society in the development and application of AI. The Lab has also launched a university outreach initiative tapping Canadian policy, legal, and computer science students who will undertake original research during the 2017-2018 academic year addressing the opportunities and challenges posed by AI for human rights.

21 September 2017

Government of China – Written evidence (AIC0145)

Thank you for your letter dated 20 July requesting information on the advances of artificial intelligence in China.

Artificial intelligence will profoundly change people's way of life and the future of the world. After years of technological exploration and accumulation, China has made considerable progress in the field of artificial intelligence, leading the world in audio and visual identification and other relevant technologies. China now ranks No. 2 in the world in terms of the number of scientific papers published internationally and patents licensed on artificial intelligence.

The Chinese Government attaches prime importance to artificial intelligence and works actively to facilitate its development in China. In July, the New Generation of Artificial Intelligence Development Plan was issued. This document aims at further enhancing China's innovative capabilities of the new generation artificial intelligence, help boost the intelligent economy and the building of a smart society. This will enable China to make greater contribution to the research and development of artificial intelligence around the world.

At present, China-UK relations enjoy a sound momentum, evidenced by effective cooperation across the board. Our bilateral cooperation in the field of artificial intelligence has huge untapped potential. I hope the enclosed materials and websites will be of some help to your committee. It is also my hope that your committee will learn more about China's progress in science and technology, so as to promote closer communication and cooperation between our two countries on science and technology-related legislation and contribute to the China-UK Golden Era.

18 August 2017

Government of Japan – Written evidence (AIC0224)

Artificial Intelligence Inquiry - responses from the Government of Japan

Please note that we have provided responses only to the questions for which we have appropriate answers.

The pace of technological change

1. "What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years?"

What factors, technical or societal, will accelerate or hinder this development?"

> In Japan, AI is expected to be used in a multitude of fields such as autonomous cars, medical diagnosis support, dialog agents, robotics, etc. AI technology is recognised as critical both for promoting economic growth and addressing social challenges such as labour shortage. The Strategic Council for AI Technology was established as a headquarter of AI technology in Japan gathering all related ministries and agencies. It is expected that technology development for AI will be intensively and extensively pursued in the future. Meanwhile, concerns about security and privacy issues require the establishment of guidelines and regulations for the AI technology. Timely and flexible adaptation of regulatory framework is critical to ensure the progress of AI technology for the benefit of society.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

> Many experts think that artificial general intelligence that sets their own objective can not be realised at the present moment or in the near future, but in the world, there is a tendency to expect too much of artificial intelligence or to worry about artificial intelligence more than necessary. When discussing AI, it is necessary to accurately grasp the realistic technical level. So far, in Japan, the government has also discussed the impact and issues of artificial intelligence on society by setting up councils, e.g. The Advisory Board on Artificial Intelligence and Human Society (CSTI).

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

> Ordinary citizens should be involved in the debate on the advancement of AI. In Japan, The Advisory Board on Artificial Intelligence and Human Society was

set up in 2016, and a report that includes a summary of the issues to be addressed regarding AI and human society was published (http://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_en.pdf). Dialogues with citizens were an integral part of the preparation of the report.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

> The benefits of AI such as improvement of production efficiency are wide, but in the short term it is considered to be limited to fields such as manufacturing industry and finance industry. It is necessary to consider ways to prevent excessive economic disparity by establishing an environment where new business can enter easily.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

> It is important for the public to understand what can be done, and what can not be done by AI technology. To that end, the government's actions are also important. As an example, in Japan, a national research and development agency, namely the New Energy and Industrial Technology Development Organization (NEDO) has created a video introducing images of the social implementation of AI, and made it public on the Internet in 2017.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

> Japan has strengths in the machine industry such as automobiles and robots. By effectively utilising big data generated in the manufacturing process and at the usage phase, these sectors may benefit from AI technology as a first mover. But the potential benefit is not limited to the manufacturing industry. We are also expecting that we can benefit from AI in a wide range of fields such as nursing care, energy, infrastructure and agriculture.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

> Many people have concerns and anxieties about AI's potential manipulation or operation of their minds and behaviour. Ethical discussions might especially be needed. Developers are expected to fulfill their accountability on the AI system they develop, so that users and society's trust in the AI system can be gained.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

3 October 2017

Government of the Republic of Korea – Written evidence (AIC0228)

The pace of technological change

Q1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

-The current state of artificial intelligence can be explained in the following ways. First, areas where AI outperforms humans: basic pattern recognition, optimization under given rules and information search. Second, areas where AI demonstrates human-level intelligence: sensation and perception. Third, areas where AI underperforms humans: inference through self-learning, logical thinking, creativity, production and understanding of natural language; and emotion recognition, inference and expression.

-The following factors have contributed to the acceleration of the development of AI: the recent development of computer hardware, advancement of algorithms such as deep learning, production of enormous amounts of data, decreased prices of sensors and development of wireless networking. Such elements of the ICT ecosystem, all together, have greatly contributed to strengthening the foundations of AI technologies.

- In the next two decades, with high performance computing, digitalization of information and ever-increasing numbers of sensor devices, AI is expected to be applied across industries (e.g. medicine, manufacturing, finance and distribution) and to many public sectors (e.g. administration, public safety, welfare and education), as well as function as the driving force that brings far-reaching innovation across the social system, society and culture.

- A large gap between developed and developing countries in the pace of technology development is expected, depending on each country's AI technology capability and underpinning ICT technologies thereof. When it comes to the impacts on society, humans are expected to lose their jobs to robots due to AI-driven automation. Countries that cannot properly respond to changes in the labor structure and job quality are likely to experience an extreme polarization in the labor force and income, exacerbating social conflict surrounding AI.

Q2. Is the current level of excitement which surrounds artificial intelligence warranted?

- The Fourth Industrial Revolution is about smarter machines improving productivity, thereby bringing a fundamental change to industry structures, and at the center of this change will be AI technologies.

- ICT-based platform companies that utilize AI technologies will expand into all

other industries, thus tearing down industry barriers and threatening existing manufacturing and service companies.

- Major countries and leading businesses are already paying attention to the disruptive impacts of AI and have been carrying out long-term and large-scale research and investment, with the growing investments and attention expected.

Impact on society

Q3. How can the general public best be prepared for more widespread use of artificial intelligence?

- The government will have to play the role of an enabler to improve the market environment in a way that encourages companies to launch AI products and services.

- To this end, the government needs to promote investment in the private sector by making public services (e.g. administration, medicine, safety, etc.) smarter.

-It also needs to build large-scale test-beds in each major area (e.g. city-based smart service, smart robot, autonomous vehicle, etc.) and make public of various data produced from these test-beds so that start-ups and SMEs can utilize them in developing new services or technologies.

-Along with such promotion strategies, the government also needs to prepare itself for potential adverse effects by implementing regulatory policies. (e.g. creating an environment of fair competition, protecting personal information, etc.).

Q4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

-In the area of AI technology, the technology gap between leaders and followers continues to grow, and those companies that dominate platforms also dominate the market. Therefore, those with platforms and ecosystems that can produce, obtain and utilize data are expected to yield the greatest profits.

- However, in the area of application service, even Micro Multi Nationals can easily launch products and services targeting global consumers by utilizing a global platform. This will provide small/new companies including start-ups with an opportunity to grow fast.

Public perception

Q5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

- AI-driven changes in industry structures will inevitably bring change to the nature of jobs and tasks, as well as every aspect of our life. Therefore, it is essential that the public understands and engages with artificial intelligence.

- In order to improve the public's understanding and engagement, the government should secure core talent in the area of AI and enhance the public's understanding of creativity and AI technology through SW/convergence education.

- The government also needs to manage budget strategically in a way to utilize AI in a preemptive manner in the public service area and all parts of society. With this, citizens should be able to actually feel and enjoy the AI-led improvement. (e.g. disease prevention, improved quality of life, reduced accidents, etc.)

Q6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? If so, how?

-To answer this question limited to industry and service, with the adoption of AI, every industry and service that we can think of will be able to enjoy increased sales and reduced costs as well as largely increased consumer benefits.

- In the case of Korea, the medicine industry is expected to enjoy the greatest benefit of new sales and reduced costs, followed by manufacturing and finance, while the transportation sector is likely to have the largest increased consumer benefits, followed by city and wellness.

Q7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

-AI technologies continue to grow in performance and sophistication through learning. Therefore, those first-mover companies that enter the market and establish an eco-system earlier than others are highly likely to monopolize the market.

- In addition, large platform-based companies can provide quality service at a lower price by collecting and accumulating data from many users; with this, they acquire more users – what is called the network effect - and generate economies of scale.

- In order to minimize the monopolistic and oligopolistic competition resulted from network effects, the government needs to create an environment where such platform companies can have fair competition. To this end, it needs to strengthen 'platform neutrality'-related systems to prevent those companies that pre-dominate services in one sector from exerting their power in other sectors through platforms.

-As AI technologies rely on large amounts of data, the government, through monitoring, should prevent few companies having a monopoly on data. It also needs to promote safe exchange of data, which have grown in quantity and quality, through measures including improved regulations on personal

information.

Ethics

Q8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

-AI advances its algorithm through self-learning. For this particular reason, developer's prejudices or existing economic inequality could sneak into the system, which could have negative impacts on the development and use of AI technologies. On top of this, one can make deadly algorithms or inject malicious data into the system (e.g. killer robots, AI viruses, etc.), raising concerns over dire consequences for humans.

-To minimize such adverse effects, the code of ethics should be established. It is to encourage ethical behavior by developers and users, thereby minimizing the risk of malfunctioning and abuse. Moreover, in the process of data collection and algorithm development, it is needed to verify the fairness and reliability of data and establish standards or procedures. (e.g. developer's obligations).

Q9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

- Unlike other software, it is difficult to understand the AI algorithm process, so uncertainty in outcome prediction is one of the important issues in this area.

- It is virtually impossible to predict outcomes just by analyzing the SW program structure. Therefore, we should be able to analyze/reason safety-related unforeseen situations, and put in place design/verification accordingly.

-In particular, in the areas that are directly related to life safety (e.g. automobile, aviation, ect.), we should be able to analyze risks that could arise when AI replaces humans in judgment, and we also need the design/verification that underpin security. However, there is no exemplary cases to be found around the globe.

- Therefore, research, under international coordination, is needed on identifying risk factors in each area; making it mandatory to have a kill switch (emergency stop) and store/manage system records; managing SW training data; and selecting verified data.

The role of the government

Q10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- Promoting the development and use of artificial intelligence requires strategic R&D investment strategies that particularly encourage universities and research institutes to promote basic science and source technology research, which are

the fundamental bases of the AI industry.

- To this end, the government needs to make a long-term investment in the areas that back theories behind AI technology including brain science and industrial mathematics. In addition, it needs to set different goals for different sectors (e.g. AI, hardware and data utilizing technology), considering how advanced technology is and how advanced technology one country has.

- Furthermore, the development of application service technologies that can be applied in the public sector (e.g. defense, public safety, welfare and culture) is needed, thus promoting the development of innovative application technologies in the private sector.

- Regarding regulations on AI, it is important to establish a flexible regulatory system for new technologies which cannot be subjected to the existing law and system. As AI technologies are something new and unpredictable, we need to consider the following possibilities: shifting a regulatory paradigm so that existing legal systems do not become obstacles, setting up clear AI development standards to prevent malfunctioning and developing deadly AI technologies, and making it mandatory to keep logging records of AI products and services that can have an impact on human's body and property.

Learning from others

Q11. What lessons can be learnt from other countries and international organizations (e.g. The European Union, the World Economic Forum) in their policy approach to artificial intelligence?

-The United States, under the lead of the White House Office of Science and Technology Policy, is developing AI technologies and devising strategies to respond to social change. Japan is also seeking economic development and solutions to address social problems with the development of AI and robotic technologies. In addition, China is actively developing technology and pursuing industrialization under the recognition that AI is the next generation growth engine.

- Considering all these examples, the government should strongly support companies in the private sector in developing AI technologies by leveraging on the UK's strengths (e.g. the AI and robot industry, etc.) At the same time, it should pursue boosting productivity through AI-driven technology innovation, exploring new industry opportunities, and addressing social problems such as low birth rates and population aging.

30 October 2017

Dr Paul Graham, Professor James Marshall, Professor Thomas Nowotny and Dr Andrew Philippides – Written evidence (AIC0088)

Dr Paul Graham, Professor James Marshall, Professor Thomas Nowotny and Dr Andrew Philippides – Written evidence (AIC0088)

[Submission to be found under Professor James Marshall](#)

Guide Dogs – Written evidence (AIC0040)

Select Committee On Artificial Intelligence

Introduction

- 1.1 Evidence submitted by John Shelton, Smart Cities Manager, on behalf of Guide Dogs.
- 1.2 Guide Dogs is recognised globally as an expert in mobility and inclusivity. For over 80 years we've been working to ensure that people living with sight loss are not excluded from life. Though best known for our guide dogs, increasingly, we are also pioneering the design and deployment of smart tech to get people out and about.
- 1.3 As the UK's leading sight loss charity specialising in mobility, we believe that everyone, regardless of their ability should be able to get out and about safely and confidently; be that to study, to work, to shop, to look after family, or to maintain their health and fitness.
- 1.4 The Oxford Dictionary defines AI as the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. The following evidence concentrates on the social application of Artificial Intelligence (AI) to inclusive smart cities, joined-up public services and customer experiences, and its positive impact on citizen wellbeing. AI has the potential to revolutionise disabled people's lives, but to achieve this inclusivity must be built in at the planning stages to ensure that AI benefits everybody.

2. The pace of technological change

- 2.1 We have come a long way since the launch of the first iPhone just 10 years ago, and the UK is leading the way on the development of smart cities; where advances in open-data and AI are finding new ways to augment and improve the world we live in. But our smart cities should not just be about the deployment of intelligent technologies to increase efficiency or economic growth; the degree to which our cities are smart also affects citizen wellbeing and the way people interact with their environments and communities in everyday life. Indeed, in 2013 the Department for Business Innovation and Skills noted that "A Smart City should enable every citizen to engage with all the services on offer, public as well as private, in a way best suited to his or her needs."⁶⁰⁷

⁶⁰⁷ Department for Business Innovation and Skills. Smart Cities Background Paper. (2013)

- 2.2 We have entered the next great era of technology innovation and thanks to advances in AI many smart city concepts now have the potential to become a reality. There is no doubting that AI has tremendous potential to benefit the UK in areas such as the economy, city efficiency and national security, but with appropriate policies, standards and funding programmes AI can also benefit our citizens at a very personal and human level, with particular potential to benefit disabled people.
- 2.3 We are witnessing a convergence between technology innovation and policy thinking; in 2017 the UK government published its long awaited UK Digital Strategy⁶⁰⁸ and a green paper setting out its plans for the UK Industrial Strategy⁶⁰⁹, and the City Standards Institute has recently published new guidelines for smart city development⁶¹⁰. Although these documents address some of the big issues and mega-trends facing society, such as energy security, transportation, population growth and urbanisation, they have not yet adequately covered the needs of the growing number of people living with disabilities in the UK.
- 2.4 In 2014 The Office for Disability Issues noted “There are over 11 million people with a limiting long term illness, impairment or disability. The prevalence of disability rises with age. Around 6% of children are disabled, 16% of working age adults and 45% of adults over State Pension age.” With so many people affected, inclusion for all sectors of society should be an important metric to attain ‘Smart City’ status, and must be better embedded within the thought leadership behind our technology and infrastructure policies, and our investment decisions.
- 2.5 Advances in AI and associated technologies, such as open data, personal robotics, autonomous vehicles, scene recognition and talking or even telepathic interfaces are just starting to revolutionise the interaction between people and their physical, virtual and mixed reality environments. Harnessing such technologies in a socially responsible manner has the potential to augment and revolutionise experience of disability and enable everyone to get around safely and confidently, accessing the information, products and services they require with minimum fuss and maximum independence.
- 2.6 The current level of excitement which surrounds artificial intelligence is definitely warranted. The following simple use case demonstrates how

⁶⁰⁸ Department for Culture, Media and Sport. Policy Paper, UK Digital Strategy. (2017)

⁶⁰⁹ Department for Business, Energy and Industrial Strategy. Green Paper, Building our Industrial Strategy. (2017)

⁶¹⁰ British Standards Institute. PAS 183 Smart cities – Guide to establishing a decision-making framework for sharing data and information services. PAS 184 Smart Cities – Guide to developing project proposals for delivering smart city solutions. (2017)

life-changing AI-enabled services can be, especially for someone living with disabilities such as sight-loss, mental health or learning difficulties.

- a. **Smart Home:** You're running late for a friend. Pulling on your coat, you simply tell the house to have the dinner in the oven cooked by 7pm, to switch the heating on at 6pm and to play some burglar-detering radio all day. A single voice activated intelligent interface will make it much easier for people with sight loss to use household appliances, avoiding the need to navigate multiple analogue technologies around the home?
- b. **Smart Streets:** You walk into town passing smart lampposts and street furniture. Their free Wi-Fi and Bluetooth ensures you are connected to intelligent services that complement your satnav so you know exactly where you are and are warned about potential hazards on route, for example that you're approaching a shared surface or temporary street works.
- c. **Autonomous Taxi:** Rain is imminent, so you simply ask your phone or wearable device to call an autonomous taxi. The integrated systems automatically know your special assistance needs, billing is fully automated and secure, and the accessible vehicle is easy to locate and use, even if you are blind.
- d. **Smart Retail:** It's your friend's birthday and you need to pick up a gift on the way. Quickly matching your budget with her tastes, your smart aisle finder easily finds the shop you want and guides you to suitable products in the store. As a blind person, smart retail takes the difficulty out of shopping as it provides you with information which would otherwise be unavailable to you, increasing independence, enjoyment and spontaneity.
- e. **Entertainment:** At your destination your tech helps you to find your friend in a crowded gallery. The customer experience doesn't just navigate you through the accessible building – the commentary empowers you to discover and select information about the exhibits on your own terms.
- f. **Public Transport:** Heading home, your tech warns about transport disruption and locates a bus stop over the road and says your bus is just 90 seconds away. Knowing exactly when your bus is due is especially helpful if you are blind, as it avoids you having to flag down every bus to check with the driver whether it is the right one. You board, aided by a driver who's received disability equality training – and you arrive safely to a warm home in time to enjoy your dinner.

2.7 The ability to deliver such joined-up experiences through AI-enabled services, that intuitively deliver the right information in the right format and at the right point and place in time, is now within reach. The individual elements of these innovations will become commonplace within the next 10 years, but greater leadership is required now from the government and its agencies to ensure that the individual elements are interoperable and operate in the whole as a seamless system. The potential benefits of this technology for people with a vision impairment are even greater than for those who are fully sighted. It is essential therefore that we do not miss the current opportunity to ensure that developers of AI and new technology ensure that their innovations are fully accessible.

3. Impact on society

3.1 Today, many everyday services and environments tend to operate in a default mode and an accessibility mode. Unless there is significant pressure during the design phase of a new system or environment, far more attention is given to the default mode, and the inclusivity mode is often not a primary concern. This tends to result in sub-optimal experiences for older people and people living with a disability, and sometimes it can mean total exclusion.

3.2 Rather than excluding millions of citizens from everyday life we need to harness AI alongside other innovations to include everybody in life, for the common good as well as that of the individual. "Social exclusion is a complex and multi-dimensional process. It involves the lack or denial of resources, rights, goods and services, and the inability to participate in the normal relationships and activities available to the majority of people in a society, whether in economic, social, cultural or political arenas. It affects both the quality of life of individuals and the equity and cohesion of society as a whole."⁶¹¹

3.3 Inclusive environments combine the design of the built environment and transport, retail, cultural and entertainment services with intelligent technologies to deliver consistent customer experiences that often transcend any single service provider's remit. Collaboration and interoperability of this nature will require central support and facilitation to seed a new approach to designing and delivering positive joined-up services, powered and enabled by appropriate infrastructures and intelligent technologies.⁶¹²

⁶¹¹ Levitas et al. The Multi-Dimensional Analysis of Social Exclusion. 2007.

⁶¹² Interoperability requires thought leadership and facilitation to seed a new approach to designing and delivering positive joined-up services, this 3-minute video aims to get the conversation started: https://youtu.be/WAC_icPov4o

- 3.4 Official disability facts and figures published on the government website in 2014 noted “A substantially higher proportion of individuals who live in families with disabled members live in poverty, compared to individuals who live in families where no one is disabled.” Although many disabled people live in poverty, the combined spending power of disabled households is still significant. The Department for Work and Pensions noted that “In 2014/15 disabled people and their families in the UK had an aggregate annual household income of £249 billion, up from £212 billion in 2012/13”⁶¹³. So, there is both a social obligation and an economic incentive to encourage innovators and service providers to embrace social inclusion.
- 3.5 The aforementioned UK Digital Strategy clearly notes the growth of digital solutions in all aspects of life and the economy, and references the need to close the digital divide. This problem has been known for many years, yet still significant numbers of the general public, particularly disabled and older people, are unable to use digital solutions because the technology and operating policies are often poorly designed, inaccessible, expensive and there is a lack of awareness and training support. This is particularly the case for people with sight loss.
- 3.6 Many disabled and older people either do not have access to the latest technology and training, or they can feel confused by the rapid pace of technological advancements. However, this sector of the community perhaps has the most to gain from advances in AI. Technology could enable people to access all the services and information they need from a single natural language interface combined with a mix of biometric security measures such as fingerprint, voice and face scanning. This could include simply asking your tech to manage bank transactions, book entertainment and transport, purchase goods, order repeat prescriptions, and so on. AI could mine, process and deliver in an appropriate format all the data relevant to an individual through a single personalised portal – removing the need for many separate apps with a multitude of passwords.
- 3.7 Whilst digital solutions incorporating AI have great potential to assist many disabled and elderly people in everyday life, it is essential to also address more traditional barriers to independence. Our future infrastructure assessments and smart city strategies also need to address the many issues associated with poor town planning, street design, building architecture and public services. In short, greater consideration needs to be given as to how advances in digital technologies and AI can work alongside physical infrastructure in the built environment.

⁶¹³ Department for Work and Pensions. The spending power of disabled people and their families. (2016)

- 3.8 The decisions and policies that we make today will affect the future, so our infrastructure policies, smart city standards, public contracting processes and economic models should proactively champion inclusivity for everyone. City administrations and solution providers must actively challenge themselves at every step to ask and answer questions such as "can this solution be used by, and be of benefit to people living with sight loss, hearing loss, or cognitive difficulties?"
- 3.9 Artificial Intelligence and smarter built environments will enable virtual and physical connectivity, they will promote independence and greatly enable disabled people to fully participate in higher education, employment and to become positive contributors to the economy. For example, research conducted in 2015 with 1,200 respondents identified that over 70% of registered blind and partially sighted people of working age are unemployed. ⁶¹⁴A shocking statistic compared to the 5% unemployment figure for the general population.

4. Ethics

- 4.1 The general public will not necessarily want to know what technology is powering their experience or who is providing the data, but they will want to know that it is secure, accurate, reliable, seamless and affordable.
- 4.2 Data protection and security are fundamental to the successful deployment and take-up of AI-enabled systems amongst the general public, particularly for disabled and older people who may be more vulnerable to negative impacts when the technology is abused or misappropriated, or during periods of technical failure.
- 4.3 Consumers will be concerned about the potential growth in spam, hacking, nuisance phone calls and invasion of privacy when AI is used for nefarious purposes. Automated systems powered by artificial intelligence are beginning to act as gatekeepers on the internet and social media, but these algorithms aren't perfect and can be manipulated for unethical purposes. AI may open a backdoor for malicious users to automatically collate personal information from many sources and to target illegal activities at vulnerable individuals, so consumers must have control over who has access to their data and why; and have the ability to prohibit access if so required.

⁶¹⁴ John Slade and Rose Edwards. My Voice 2015: The views and experiences of blind and partially sighted people in the UK. RNIB. (2015)

- 4.4 There is also a concern that some public services and life-affecting decisions may become de-sensitised due to an over-reliance on AI to perform functions that traditionally require a human touch.

5. The role of the Government

- 5.1 Undoubtedly, we are on the cusp of multiple technologies converging at a time when government support and guidance is most needed to ensure that marginalised communities are finally brought into the mainstream in terms of education, employment, health management, social wellbeing, employment and net contribution to the economy.
- 5.2 Interoperability is equally as important between geographic areas as it is between technological systems. In a climate of increasing devolution, city administrations require clear guidance and support to ensure that their smart cities are actively focussed on delivering inclusive environments and communities for disabled people and their families. 'Inclusion' must be given equal prominence and investment amongst other smart city themes and the government should proactively challenge city administrations and technologists to identify and implement AI-enabled solutions to ensure that no-one is left out of life.
- 5.3 There are many research projects underway in the UK and abroad investigating how advances in technology can assist disabled people, but these projects require government support and leadership to access real-world infrastructure to break out of the research environment to become scalable and interoperable solutions.
- 5.4 Strong proactive steps are needed now to design-in inclusivity on all new technology and infrastructure projects - be they schemes to replace old street lights with intelligent lampposts, or the roll-out of autonomous vehicles from 2020 onwards.
- 5.5 If the government does not actively focus attention now on 'inclusivity' at all stages of the strategy, design and implementation processes, then the UK will fail in its drive to create smart cities – as in cities that continue to exclude millions of citizens cannot be considered smart!

31 August 2017

Dr Ozlem Gurses, Dr Antonios Kouroutakis, Dr Valentina Rita Scotti, Dr Aysegul Bugra and Matthew Channon – Written evidence (AIC0051)

Dr Ozlem Gurses, Dr Antonios Kouroutakis, Dr Valentina Rita Scotti, Dr Aysegul Bugra and Matthew Channon – Written evidence (AIC0051)

[Submission to be found under Dr Aysegul Bugra](#)

Baroness Harding of Winscombe – Written evidence (AIC0072)

From DIDO HARDING, BARONESS HARDING OF WINSCOMBE

INTRODUCTION – WHY WE SHOULD HAVE THIS DISCUSSION NOW

- 1 I would like to submit evidence to the Select Committee on Artificial Intelligence specifically regarding public perception (question 5) and the ethical questions (questions 8 and 9) that the committee poses.
- 2 Firstly, I'd like to say how encouraging it is to see the House investigating this issue. The growth of artificial intelligence will present tremendous opportunities for society but this technology will also create significant opportunity for great harm to be done. As Sir Tim Berners Lee famously said in 2009: "The Web as I envisaged it, we have not seen it yet, the future is so much bigger than the past"
- 3 We have (or should have) learned a few key lessons about the digital world by now: Firstly, It moves faster than we ever think it will; Secondly, we can't always predict where it will go next; And finally, it is a morally neutral thing, it takes on its morality from what we put into it. I expect AI will follow all these rules, to the power of ten.
- 4 And it will have social and political consequences as well as economic ones. If machines will over time be capable of doing many of the roles currently done by humans (some estimate as many as 45% of current jobs will be automated over the next 20 years), we not only need to retrain the workforce, but, just as with the industrial revolution, we need to make sure that we debate and discuss how to accrue the benefits to society and how best to mitigate any potential downsides.
- 5 There are great examples where we have had the moral and ethical debate as technology was developing and the possibilities the technology offered began to emerge: the most often quoted is the Warnock Commission that reviewed human fertilisation and embryology in the 1980s, which settled public opinion, set the framework for balanced regulation in the UK and enabled the UK to benefit – citizens and businesses- from the development of the technology.
- 6 There have also been examples in the UK where we have not had that debate. Arguably the best example here is GM crops, on which British opinion is still very much divided, our regulatory approach is patchy and much of the technology development is happening elsewhere, whether we like it or not.

- 7 There are voices who will tell the Committee that it is too early to have the ethical debate (mainly from those who would prefer no regulation for as long as possible), but I would argue that we cannot afford to delay the discussion. My assumption is that in time, the combination of the internet of things⁶¹⁵ and machine learning⁶¹⁶ will bring enormous opportunities to predict human behaviour, and therefore to improve health outcomes and to enhance life experiences. It will also bring about the opportunity to manipulate human behaviour for malicious as well as benign purpose and create seismic shocks to the economies of the world.
- 8 Working out how to create an environment which encourages the innovation this technology can bring, but also has a safety net that protects the vulnerable in society is not simple and needs time to be discussed and debated.
- 9 And we may not have very much time, as the technology is moving at pace. Indeed, we will probably never be able to keep pace with the technology once it properly emerges. We need to get ahead now, as fast as possible.
- 10 Already today we have companies like Ocado trialling driverless vehicles to deliver groceries; others using AI to conduct initial Skype interviews⁶¹⁷; traders using AI to allocate block trades; Prison authorities using AI to predict likelihood of reoffending and therefore whether to grant parole.
- 11 It's entirely possible that these applications improve the fairness and accuracy of the processes, are more efficient and more effective. It's also entirely possible that the training data used to create these AI models is not fair, has inherent biases based on the inputs and goes unmanaged and unchallenged.

PUBLIC PERCEPTION

- 12 People are often scared of things they do not understand. And almost all of us don't understand how AI does, and will, work. Arguably one of the worst outcomes could be the UK suppressing AI because people are afraid of it. There are many examples of this happening over the centuries, from printing presses being banned in the Ottoman Empire (leading to dramatically lower literacy rates than in Western Europe for several hundred years) to German boatmen sinking the first steam powered ship. In virtually all these examples no one succeeded in stopping the march of technology, they just ensured that someone else benefitted.

⁶¹⁵ By which I mean the ability to build very low cost monitoring/data capture devices into almost everything we use

⁶¹⁶ The ability for machines to analyse the huge amounts of data that will be created by the Internet of things at speeds and in ways that a human brain could not process

⁶¹⁷ Financial Times Series "Artificial Intelligence in Real Workplaces"

- 13 Britain's openness to new technology and industrial innovation was one of the key reasons that we emerged as one of the most prosperous and civilised nations in the 19th Century.
- 14 We are, I fear, already seeing some of the early signs of societal discomfort with the rapid pace of change at least in part driven by technology change. At a macros level segments of Brexit voters were undoubtedly driven by a desire to slow down societal change and shouting loudly that globalisation and the associated technology revolution wasn't working for them. Rather than dismiss this, I think we need to hear the concern and recognise it as very real. The rapid pace of technological change may be only one of those fears, but it is only going to get bigger if we don't address it.
- 15 So, I believe that one of the most important things we need to do as a nation is increase the level of public understanding of the digital world in general, and AI as a matter of particular future importance. In this way, we will be preparing people across society for the changes that lie ahead, and enabling citizens to feel that they are genuinely in control of the technology rather than the other way around.
- 16 This is not just about teaching coding to school children. We need to think much more comprehensively about how we build:
 - a) public confidence in technology, including AI and VR
 - b) basic digital understanding
 - c) general technology skills that will be required by everyone in the workforce
 - d) specific technical expertise to develop AI itself.
- 17 Let's take a small example that is not far away: driverless cars. It's quite probable that if properly developed and tested, driverless cars will be much safer than human drivers. But it's highly unlikely that most people will think so initially. Public confidence will come from a combination of education, but also sensible ex ante regulation. We don't want the equivalent of the man with a red flag in front of the driverless car, but we also don't want cars being tested on the road without sensible constraints. And society needs to be prepared to debate the obvious ethical issues that need to be resolved to ensure that the driverless cars makes decisions that we view as acceptable when faced with bad vs bad trade-offs (do I run over the granny or crash into the tree?). Not least because its also entirely believable that long before driverless cars are mass market, they will be overtaken by driverless flying vehicles, which may make us all feel that this is now part of sci-fi movie, but is real enough that the issues need to be discussed before the cars are in the air not after.
- 18 We also need an adult population able to manage the technology. It's not going to be acceptable to delegate managing the driverless car to your

children, in the way our parents did for VCRs, or, for example, to handover the development of the highway code solely to the software engineers developing the cars. Everyone will need a basic understanding of how the technology works to effectively integrate it into society. And finally, we want Britain to be a country whose software engineers are at the forefront of the development of the technology itself, so our education system also needs to train and develop real experts.

- 19 I would encourage the committee to take evidence from experts in all four of these areas rather than just one. We will improve public understanding and the country's capability to use the technology if we have better primary, secondary, tertiary and on the job technical education. But we will also need to lead a debate about ethics and regulation – just as Dame Mary Warnock did – in advance of the technology scaling if we are to ensure that society is ready for the difficult ethical decisions ahead.

20 ETHICAL ISSUES

- 21 This leads me to the ethical issues that AI could present. I agree with the Royal Society in their report on Machine Learning that it is essential that "Society needs to give urgent consideration to the ways in which the benefits from machine learning can be shared across society". To do that effectively, I think we should be constituting an independent expert body to scope, debate and recommend approaches to manage the potential ethical issues sooner rather than later. I think the broad ethical issues of the growth of AI are huge. Here are five areas to consider which I am sure are not comprehensive:

- 22 **1. How does a civilised society manage the displacement of jobs that the rise of AI will create?** There is no doubt that AI is doing to make millions of human jobs redundant. It's also highly likely that technology will also create many millions of new jobs. Whilst economists argue about which will be the larger number, what I think is certain is that there will not be a seamless transition either in the volume of jobs lost and created, the skills required vs the skills being developed at any level of local or national geography. So those of us that believe in the benefits of a socially responsible, democratically overseen market economy need to think hard about how we manage the transition that AI is undoubtedly going to bring so that communities across the country genuinely benefit from the rise of AI. Arguably we are already way behind in addressing this.

- 23 **2. How should ownership of the machines and the wealth they will create be regulated?** Already there is some debate, much like that led by Marx in the 19th century that Machines should be owned by the people for the people. Again, if you believe in a the benefits of a capitalist, market economy with a democratic safety net, we need to start working through where the societal safety net needs to be extended to regulate AI. And this needs to be

done in a pragmatic and collaborative way. I understand that the majority of Amazon's drone testing worldwide is currently being done in the UK because Amazon is finding that the UK regulators are willing to work with them, upfront to develop the regulations alongside the testing. This has to be good for the UK and just public/private; regulator/innovator collaboration is going to be essential across a range of sectors.

- 24 **3. How do we govern the use of the data that AI will rely on?** How do we ensure that the digital world is both Open and Safe (as the Royal Society puts it)? What rights should an individual have to their own data versus their responsibilities to society in general to share their data if sharing can bring great societal benefits in e.g. healthcare? How do we ensure that individuals' data rights are protected? . This debate is beginning under the auspices of translating the GDPR into UK law, but there are deep seated ethical issues about data rights that need to be addressed for the long term if individuals are to have confidence in a modern digital society.
- 25 **4. How, if at all, should we constrain the use of AI? Are there things that can or will be done, that should not be allowed?** Nearly 50 years ago the UK decided to ban the use of subliminal advertising. Are there forms of highly targeted data based marketing that are far more effective than subliminal advertising that should be banned? Or are there types of campaigns that would not be possible without AI that should be banned? E.g. should it be legal to run voter suppression campaigns that are individually targeted to persuade you not to bother voting? Equally on the other side, how do we ensure that innovation isn't constrained by vested interests who currently control the old technologies and benefit from slower rollout of the new. There is much less discussion about data usage than there is regarding Data Governance, and yet in the end it is what is done with data that can enhance or damage society.
- 26 **5. How do we audit AI? Those who build them, those who manage them, those who use them for good or ill?** In addition to be scared of the unknown, there is also a risk that society is too accepting of experts and the technology itself. How do we create the right audit tools to ensure that the algorithms are not breaking the law? For example, racial profiling is illegal. How do you ensure that an algorithm being used to predict reoffending rates, that is commercially confidential, isn't racially profiling? And once the algorithm is being created by the machine, the audit task becomes even more difficult. Some of the most pernicious forms of bias are unconscious, not conscious bias. How do we audit AI for unconscious bias? How do we embed the principles of ethical design into software engineering? And then audit compliance?
- 27 150 years ago, health and safety legislation began to set out what was and was not required of factory owners. Over the last century the standards have

developed, training and development of factory management has changed hugely and so has the audit and compliance functions of the state. We need to develop the ethical framework for machine data usage and what should and shouldn't be expected of the AI's human managers, and build the societal capability to audit that framework.

- 28 In a short submission such as this it is impossible to do just to the subject of data ethics, save to say that there are so many important ways that unethical application of AI could change society for the ill and how ethical applications could transform the world for the good, that I think it is critical that we start the debate and discussion as soon as possible.
- 29 And whilst parliament has a huge role in initiating the discussion, it's important that we take that away from the short term political debate and enable technologists, philosophers, lawyers and lawmakers to debate and discuss these issues through an independent body or Royal Commission with a wide ranging brief and an ambition to enable Britain to lead the world in creating a civilised digital country that makes the benefits of technology work for everyone.

4 September 2017

HM Government – Written evidence (AIC0229)

GOVERNMENT RESPONSE TO WIDER CALL FOR EVIDENCE FROM THE HOUSE OF LORDS COMMITTEE ON AI

FROM THE DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT, and DEPARTMENT FOR BUSINESS ENERGY AND INDUSTRIAL STRATEGY

The pace of technological change

Artificial Intelligence could be described as the simulation of intelligent behaviour by machines, either through programming or by learning themselves. The field of AI has evolved in several waves across research domains, and is now broadly recognised as a group of technologies based on the development and application of computer algorithms based on statistical methods.

Data and AI technologies are now permeating our lives at an ever-increasing rate, from smart features built into mobile operating systems that autocomplete our messages and identify meeting schedules and venues in our messages, to voice-activated smart home assistants and the ability to tailor our transport itineraries.

The advent of what is sometimes referred to as the “fourth industrial revolution”, sees AI technologies and intelligent automation entering into many industries, such as manufacturing, utilities and raw materials extraction and processing. This often combines technologies such as additive manufacturing, Internet of Things, virtual and augmented reality, and biotechnology.

Computer-based Artificial Intelligence has a long history, dating back to Alan Turing’s research after World War II. Since then, research and development has evolved through surges of research activity being conducted across many disparate fields. The last decade has seen an unprecedented pace of development, as many of the fields which form the foundation of Artificial Intelligence, such as computer science, systems thinking and bayesian statistics, have matured rapidly. These have converged with the proliferation of digitised data and advances in processing power and communication bandwidth to make these technologies economically viable in many consumer applications.

The current wave of consumer AI technologies are largely built on the benefits of intelligent automation and autonomous systems. These have been driven forward by companies successful in key areas such as online search, online retail and digital social interaction (messaging and content sharing). These services rely on access to free flows of data – indeed this is a key characteristic of the digital economy.

The increased accessibility of the internet; connected sensors, devices and machines (the Internet of Things); and exponential increases in computing capacity and availability will continue to be critical factors in the rapid development of Artificial Intelligence.

Improvements in data processing throughput, expansion of warehouse data centres, allowing cloud computing and connectivity, and fast processing of large quantities of data have also been catalysts. Artificial Intelligence in general and machine learning in particular is dependent on the flow and availability of good data sources to train software at speed and scale not previously thought possible. Many recent successful applications of AI have been made by businesses that apply these data processing capabilities on an industrial scale to develop algorithms from machine learning programmes.

If the pace of recent developments is indicative of the pace of change ahead, this presents the UK with a really significant opportunity, as it has the fundamental strengths to be a major player in the artificial intelligence market.

Impacts

Alongside the realisation of the technological potential, Artificial Intelligence, machine learning and automated production have the potential to change our world faster and more fundamentally than any previous technological revolution.

The debate about whether AI is overhyped is evolving into a debate about how to harness the potential, maximise the benefits responsibly, and distribute them equitably.

Impacts in industry, for example manufacturing, raw materials extraction and processing, and utilities, are likely to be profound in terms of productivity, the nature of work and jobs, and the skills required to drive this transformation.

We have a resilient and diverse labour market in the UK, demonstrated by the latest record-breaking figures showing more people in work than ever before. Whether in cyberspace or on the shop floor, advances in technology bring new jobs. It is only right that we embrace these opportunities, support new skills and help more people get into employment to secure a workforce of the future.

Jürgen Maier's Review of Industrial Digitalisation, the recently published *Made Smarter*⁶¹⁸ report considers more widely how we the Government can work with industry to ensure the benefits of new technologies are felt in different sectors of the economy, creating new, exciting and well paid jobs across the country.

Ethics

⁶¹⁸ <https://www.gov.uk/government/publications/made-smarter-review>

The use and availability of data is helping businesses, the public sector and citizens in many positive ways through the application of AI

We need to ensure that organisations, citizens and Government can use data to make decisions and provide goods and services and people's lives, and ultimately improve the economy.

This means ensuring that we have the policies and regulatory structures we need to respond quickly and intelligently to new developments in uses of data, and that those making decisions about the way data and technology is used do so in an informed and ethical manner.

The Conservative Party 2017 Manifesto committed to setting up a new body to advise Government and regulators on the ethical use of data, which AI applications depend upon. The body will develop an effective ethical framework to help govern the use of data, and the impacts of decisions made from that data.

In 2016, Government published an ethical framework for the use of data science within government, which is currently being updated. This framework will ensure responsible application with accountability and fairness in the use of data technologies across government, with accountability and fairness, and could also be useful for other organisations.

Industry

Government's Industrial Strategy Green Paper (January 2017), and Digital Strategy (February 2017) identified AI as a major, high-potential opportunity for the UK to build a world-leading future sector of our economy. Government is encouraging uptake of digital technologies for economic and social benefit across industries and the public sector. Helping every business be a digital business is a vital part of increasing the productivity of UK business. In 2016, overall tech investment in the UK was £6.8bn – more than twice any other European country.⁶¹⁹

It is important to ensure that organisations and citizens can take advantage of the opportunities. We recognise the assessment in the AI Review that AI creates opportunities for all sectors of the economy. It appears to be one of the fastest growing segments of the digital sector. In the UK, Tech City UK estimated that AI received 3% of the investment in digital tech in 2016, and was rising. But the benefits of AI are not restricted to a single industry. We recognise that data

⁶¹⁹August 2017: <http://www.telegraph.co.uk/business/open-economy/why-uk-should-top-world-in-tech/>

science and Artificial Intelligence technologies offer opportunities across all sectors.

The responsible and innovative application of AI can unlock the power of data for the UK economy and bring great public benefit. Many sectors across the UK economy are already embracing innovation through AI and benefitting from its use in how they do their day-to-day business.

All the major global tech companies active in the UK are developing and using AI, and there is a healthy start up ecosystem for companies focussed on AI technologies. The UK has produced a number of very innovative AI companies, and companies are being formed frequently. According to one report, a new AI startup has been founded in the UK on almost a weekly basis in the past 36 months. Another study has counted 226 independent, early stage AI companies in the UK, which is almost double the number of companies in the second highest European nation.

Government and the public sector also stands to benefit from AI, particularly in the delivery of services. AI chatbots can assist with routine calls from users of local government services, for example Enfield Council's much-publicised trial of a chatbot in its planning department. Defra's Earth Observation Centre of Excellence is exploring using AI to help process and analyse satellite imagery to identify changes in the landscape during and after environmental incidents. The cross-government Centre for Controlled and Autonomous Vehicles (CCAV), is jointly run and resourced by DfT and BEIS and heralded by government as a successful model of cross-Departmental work. CCAV is undertaking a programme of work that covers policy, strategy, and research and development, in order to ensure implications around the operation of autonomous vehicles can be fully considered, and their benefits enhanced, in advance of their delivery to market in the coming years.

The AI Review has highlighted the UK's existing strengths in AI research and business, and the opportunities for further growth; the supply of skills and talent needed to fuel the industry; access to data, in particular for small and medium companies; how start-ups, SMEs and universities are best able to collaborate and interact with larger tech and AI-focused companies and how a new forum will help lead and coordinate this nascent industry.

Further work is being taken forward by DCMS and BEIS for the Industrial Strategy White Paper.

The Role of Government

Government is committed to helping industry innovate and advance, by shaping the right opportunities in skills, investment, governance, innovation and business support to foster the right environment for UK AI companies to develop and scale.

Technology is developing faster than society can develop ways to deal with the challenges it creates.

Government can, for example, support people and organisations with high quality advice and guidance, underpinned by regulation and non regulatory action, for example by creating incentives.

Government cannot achieve its ambitions working alone. We must work with industry, across the public sector and society, and indeed other governments and international fora, for the benefit of the UK economy and citizens.

A strong framework and regulation exists to address data protection and privacy, in the form of General Data Protection Regulation and the Information Commissioner's Office (ICO). The ICO also works closely with other regulators such as the Financial Conduct Authority (FCA) and Competition and Markets Authority (CMA). AI will create new challenges for regulation in the future, and it is important for all sector regulators to be part of the adaption of systems where required. The planned data and AI ethics body will support this by working closely with regulators and stakeholders.

Government has made clear its commitment to digital connectivity in the Autumn Statement 2016 and Digital Strategy. This future infrastructure will be needed for consumer and citizens to take advantage of AI services. As part of a £1bn package of announcements made to boost the UK's digital infrastructure, we are funding a coordinated programme of integrated fibre and 5G trials to ensure that the UK leads the world in 5G connectivity.

Government also has a role to ensure that adults who lack core digital skills can access basic digital skills and training where it is available. We have legislated for this in the Digital Economy Act, and DCMS are now developing the detail of the policy with DfE and BEIS.

In July this year, the Digital Skills Partnership was launched, bringing greater coherence to the provision of digital skills training at a national level, and supporting the development of local level partnerships to increase the digital capability needed to build inclusive, thriving local economies. It brings together partners from public, private and charity sectors to collaborate on this very important agenda. The first board meeting for the Digital Skills Partnership will take place in November.

Other Stakeholders

Today the UK is a world-leader in the science underpinning this technology, with a rich ecosystem of investors, employers, developers and clients, and a network of supporting bodies.

Our universities are a major source of the top talent in the field, and several of the world's most innovative AI companies are based here, within a rich tech sector, represented by TechUK.

The Alan Turing Institute is at the heart of the knowledge quarter, and has already responded positively to one of the AI Review recommendations suggesting that it becomes the national institute for Artificial Intelligence.

The Royal Society and British Academy have undertaken important work on Machine Learning and Data Governance, also working with government officials.

The Government will continue to work with these colleagues and many others in the coming months to refine the ways in which we can ensure the benefits of AI can be realised and distributed

ANSWERS TO SPECIFIC QUESTIONS IN YOUR LETTER, FROM THE MINISTER OF STATE FOR DIGITAL, The Rt Hon Matt Hancock MP

1. Has the Government defined Artificial Intelligence? If so what is that definition?

In the report by the House of Commons Science and Technology Committee on Robotics and Artificial Intelligence in September 2016, they stated that there is no single, agreed definition of Artificial Intelligence.

A GO-Science paper in 2016, Artificial Intelligence: a Guide for Policy Makers, concurred with the above, noting there are many “different definitions of ‘Artificial Intelligence’, ‘machine learning’ and related terms”, and that ‘Artificial Intelligence’ is a broad term which more generally refers to “the analysis of data to model some aspect of the world, and provide inferences from these models.”

The EPSRC (Engineering and Physical Science Research Council) has a widely recognised definition which represents the research definition which builds upon Alan Turing’s original concept of AI: “Artificial Intelligence technologies aim to reproduce or surpass abilities (in computational systems) that would require ‘intelligence’ if humans were to perform them. These include: learning and adaptation; sensory understanding and interaction; reasoning and planning; optimisation of procedures and parameters; autonomy; creativity; and extracting knowledge and predictions from large, diverse digital data.”

The Information Commissioner’s Office (ICO) states that the terms ‘big data’, ‘AI’ and ‘machine learning’ are often used interchangeably, but there are subtle differences between the concepts. They distinguish AI programs as ones that learn from the data in order to respond intelligently to new data and adapt their outputs accordingly.

We acknowledge that there are many distinctions between specific technologies and terms such as Artificial Intelligence, machine intelligence, deep learning, machine learning.

These descriptions share a common recognition of the interchangeability of terms within a broad group of technologies, which have all been rapidly developing with the increase in computing power and the availability of data.

We therefore recognise the usefulness of the umbrella term used in the Independent Review of AI published on the 15 October, as a group of complementary general purpose digital technologies enabling machine to do complex tasks.

2. How is AI defined when classifying businesses operating within the UK?

Contributions of the AI sector are captured in official statistics on the economy such as Gross Domestic Product (GDP) and Labour Market Statistics, although the contribution of the AI sector is not fully separable from the rest of the economy.

There are challenges in working with Office of National Statistics (ONS) data to identify businesses involved in Artificial Intelligence – including the design, production or sale of AI goods and services under the ONS Standard Industrial Classification (SIC) 2007, used by producers of official and other statistics in the UK, as there is no dedicated SIC code for this type of activity. This is a common problem for new technologies and sectors such as AI, Cyber Security, and the Internet of Things – but elements of all of these sectors are identifiable in electronics and manufacturing codes. Government continues to work with the ONS to better measure and reflect the UK's strengths in new technology.

Beyond ONS statistics, there are a range of potential approaches in terms of data sources and methods, which could be used to classify or measure the UK AI sector in the future. Big data may be a particularly relevant source of information to determine the size of the AI sector as its growth is inextricably linked to the evolution of the internet and technology. The use of techniques such as web scraping offer some value as many (if not all) AI businesses are likely to operate through a website and so technologies could be used to identify and classify businesses in the AI sector through keywords and phrases. The Digital Catapult and Open Data Institute have, for example, sought to analyse the UK IoT sector and "ecosystem" using experimental techniques. There are important methodological, legal, ethical and governance issues to be worked through with these kind of approaches.

3. What AI tools and programmes are the Government using or looking to use in the near future (in both cross departmental work, and within departments and agencies, including the military and security services)?

Government departments and agencies have been building their data science capacities over the last few years, with the aid of the Government Data Science Partnership, consisting of the Office of National Statistics, the Government Digital Service, and the Government Office for Science. As this capacity has developed, departments and agencies have started to develop a wide range of machine learning and AI applications. The list below is representative of the many uses to which AI is being applied.

The Government Digital Service (GDS) uses machine learning to help automate and process user comments from surveys on gov.uk as well as predicting peak traffic demands to the most popular content searched for by the public.

GDS works with the Pensions Regulator to improve efficiency using predictive algorithms for future pension scheme behaviour and HMRC uses AI to help identify call centre priorities.

There are plans to experiment with more comprehensive machine learning applications across Government. The ONS Data Science Campus, launched earlier this year at the Office for National Statistics acts as a hub, bringing together data and digital expertise and leadership across government. The Campus aims to gain practical advantage from the increased investment in data science capability, and help cement the UK's reputation as an international leader in this field.

The Data Science Accelerator programme and Government Data Science Partnership train data scientists across government in advanced data analytics including machine learning techniques to gain new insights into live departmental services and processes. Knowledge and best practice relating to use of AI to tackle policy and operational challenges are widely shared across the government data science community through conferences such as the recent, inaugural Government Data Science Conference. The Office for National Statistics provides an MSc in Data Analytics for Government and Data Science apprenticeships.

The Government's Data Advisory Board works to align cross-Government efforts to leverage the potential of data science in departments, with a particular focus on its value as an input to broader policy making processes.

The Cabinet Office Government Commercial Function is running a procurement process to bring in a strategic partner to help promote Robotic Process Automation (RPA) and accelerate uptake. This will be a vehicle through which Departments can identify, develop and purchase RPA solutions. The Cabinet Office is also exploring how central commercial arrangements and government standards could support use of robots and AI. There are also existing channels for collaboration on innovation in addressing public needs, notably the Small Business Research Initiative (SBRI).

HMRC have been engaging in a programme looking at the possibilities of using AI and machine learning to improve their processes and services. HMRC have started to use it for contact handling, casework decision making, and helping customers through effective self service, as part of a goal to automate 10 million processes by the end of 2018.

AI has many potential leading edge applications within a defence context, and will have a significant role in delivering efficiencies and shaping future operational and competitive advantage. We recognise that AI is a developing field that has the potential to further transform how defence operates. The Ministry of Defence (MOD) is committed to continuing to invest in this innovative,

emerging area and support research and development to retain our technological advantage, including through the Defence Science and Technology Laboratory (Dstl). Defence also has an £800m Innovation Fund to provide the freedom to pursue and deliver innovative solutions, with a first challenge to revolutionise the human-information relationship for Defence.[1] Current examples of AI initiatives in defence include:

The Royal Navy has established Project NELSON as a centre of excellence in Data Science and AI, aiming to put a Royal Navy owned Artificial Intelligence at the centre of its warships.[2]

The British Army has established a Capability Spotlight[3] to focus on the opportunities offered by rapid adoption of Remote and Autonomous Systems (RAS), including AI. It has also sponsored the Last Mile Challenge[4] for autonomous resupply. One of the 3 strands of this challenge is to use AI/Machine learning to reduce the logistic load on the front line by more intelligently forecasting demand.

The Defence and Security Accelerator publishes innovation competitions online, which include work on AI. Some relevant examples include revolutionising the human interface with information[5] and the challenges to allow rapid and automated integration of new sensors[6], free up personnel through innovative use of AI[7], and make effective use of human-machine teaming.[8] A MOD AI Hackathon is planned for November 2017[9].

The MOD has adopted an innovative approach to engaging with a broader supplier base for data science solutions through both a public data science competition on 'Kaggle'[10], and our own Data Science Challenge Platform[11]. Dstl also fund research into how automation and machine intelligence can analyse data to enhance decision making in the Defence and security sectors[12].

MOD is a key partner in a strategic relationship with GCHQ and the Alan Turing Institute[13], to access scientific and technological advice, accelerate applications to defence and security projects, and develop the most advanced data science skills within our analytical professions and scientific community, continuing to build on the legacy of Alan Turing.

4. What is the outcome of the Hall-Pesenti Review? How will it be used to inform government policy? Will a copy of the report of the review be published?

As the Committee will know, the final report of the independent review, '*Growing the Artificial Intelligence Industry in the UK*', was published on 15 October. The proposals in that review set out by Dame Wendy and Jerome Pesenti make 18 recommendations for how to make the UK the best place in the world for businesses developing AI to start, grow, and thrive.

Government has welcomed the report and its suggestions for how Government can work with industry to stay ahead of the competition and grow the UK's use of AI right across the economy.

Government is taking steps through the Industrial Strategy and Digital Strategy to help ensure that the UK maximises the wider use of digital technology to increase productivity, and create high-skilled jobs. We aim to make the UK the best place in the world to establish and grow a tech business; support technologies from the laboratory right through to the marketplace; and to have a supportive regulatory regime that fosters innovation in tech.

Both the Industrial Strategy and the Digital Strategy identified AI as a major, high-potential opportunity for the UK to build a world-leading future sector of our economy.

As the BEIS Secretary of State has indicated in his initial comment on the Review, work is now underway to negotiate an ambitious AI Sector Deal that will include specific measures, informed by the Hall & Pesenti review, to realise this vision. Further to the answer to question 3, above, this will include consideration of recommendations around increasing the application of AI in central Government, local government and the wider delivery of public services.

Government officials will continue working alongside the authors, and others with an interest, on actions to make the UK a world leader in this transformative new technology. We will also continue to work across sectors and the nascent AI sector in the coming months to secure an ambitious Sector Deal which helps us harness the opportunities and make the UK a world leader in AI development and the positive adoption of AI technologies.

5. What previous Artificial Intelligence-related initiatives, government-sponsored research and development, or wider Government Policies have there been towards AI?

The UK is a world-leader in the science underpinning this technology, with a rich ecosystem of investors, employers, developers and clients, and a network of supporting bodies. Our universities are a major source of the top talent in the field, and several of the world's most innovative AI companies are based here. The Royal Society and British Academy have already undertaken important work on Machine Learning and Data Governance, working closely with the government officials.

Innovate UK run a number of programmes across the areas of Big Data and Artificial Intelligence, some of which are detailed below. These are also focus areas for the Digital Catapult.

Following the 2016 House of Commons Science and Technology Committee's report on Robotics and Autonomous systems, a Robotics and AI Special Interest Group was established through Innovate UK's Knowledge Transfer Partnership enabling UK RAI innovators to connect, showcase their capability and access markets, both at home and globally. The robotics industry is also beginning a dialogue with government around a Sector Deal

£16m has been made available in the first wave of the Industrial Strategy Challenge Fund for Robotics and AI demonstrators and collaborative R&D projects. Government has committed to investing a total of £93 million in this challenge area over the next 4 years.

As outlined above, the AI Review sets out how Government can work with industry to stay ahead of the competition and grow the UK's use of AI right across the economy.

The Government is considering how to develop the 2017 Manifesto commitment for a Data Use and Ethics Commission. This body will advise the Government on the measures that are needed to create an effective, ethical framework for innovation in data use and AI technologies.

Government also has an important role to play in ensuring that our workforce is equipped to respond and is taking actions at all stages of the digital skills pipeline. We have introduced a new computer science curriculum that focuses on computational thinking and problem solving, and innovative digital degree apprenticeships and we are reforming technical education, including creating a specialist digital route with a clear pathway to employment.

In the March Budget 2017 we announced spending of up to £40 million to test different approaches to help people to retrain and upskill throughout their working lives.

2 November 2017

Fabia Howard-Smith and Laurence Freeman – Written evidence (AIC0147)

Authors: Laurence Freeman and Fabia Howard-Smith work within a London based technology corporation. Laurence’s degree in AI combined with Fabia’s personal interest spurred them to complete the responses and answer the following questions. These views are our own.

1. Introduction

1.1 Introduction

Over centuries of evolution, natural selection has gifted the human population with a strong capability for pattern recognition. Pattern recognition and its overarching ties with humanity⁶²⁰, is a topic that has been greatly researched by the field of both neuroscience and artificial intelligence. If we consider pattern recognition from an anthropological standpoint, individuals with unique pattern recognition skills are rewarded by society; from the hunter-gatherer period to the modern day, an individual’s place in society was partly founded on their ability to hunt, or more recently, to predict trends in the financial markets or even to spot a common cold in a patient. Conversely, from a technology perspective, we look to intelligent software to identify patterns for the purposes of automating routine labour, understanding speech or images and even for diagnostic classifications in medicine, all of which, are solved by varying degrees of AI.

1.2 Definition

Artificial intelligence (AI) can be described in a most simplistic form: pattern recognition software. At first it might seem that such a definition is overly simplistic, and in order to illustrate the definition further the reader will find examples of AI throughout. But in order to be more concise, one can define artificial intelligence as “The use of mathematical models to analyse and predict the patterns present in nature”.

2. The pace of technological change

2.1.1 What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

2.1.2 Current state of artificial intelligence

Within the last two decades, computer scientists have used AI for the purpose of designing: autonomous cars (computer vision), virtual assistants (natural language processing & voice recognition) and medical diagnosis machines

⁶²⁰ Ripley, B. D. (2007). *Pattern recognition and neural networks*. Cambridge university press.

(computer vision) etc. Recent developments in the aforementioned fields (e.g.: computer vision) are partly due to the arrival of large open source data sets such as ImageNet, that allow researchers access to over 14 million pictures. The social movement to make scientific findings more transparent is being achieved by an open data-shift⁶²¹; international bodies are following a similar trend towards sharing data, such as the ESA's (European Space Agency) open access policies.

2.1.3 How artificial intelligence is likely to develop

Over the next 20 years, large scale data availability and processing power will continue to be the driving factors underpinning the success of AI. The exponential growth that follows technology, stated in Moore's law, will allow AI scientists increasing access to processing power which lays the foundation for AI software to "learn" faster. And as we sit on the precipice of the quantum computing era (which promises an exponential increase in computing power), over the next 10 years, society will see more examples of general AI. This forecast is derived by the recent investments in quantum computing by Lockheed Martin (The largest US military weapons corporation), and the setup of QuAIL (Quantum Artificial Intelligence Lab) by the US government (NASA), partnered with Google.

2.1.4 Factors that have accelerated this development

Recent increases in computational power, combined with general purpose algorithms, have elucidated previously unsolved mathematical problems/objectives. For example, Google DeepMind's AlphaGo AI software managed to beat the world Go (an ancient Chinese game) champion – a task previously thought by scientists to be decades away from achievement due to Go's intuitive nature. Such a feat was made possible by the rise of processing units tailored for AI (Google's TPU (Tensor Processing Unit) is an example). Additionally, technological advancements have led to the development of Chatbots using AI, which are increasingly improving at modelling human behaviour; creating a huge impact for businesses and in particular the customer service industry. It can be considered that we're approaching a point in time where humans and bots are no longer distinguishable. In 2014, a bot named Eugene Goostman⁶²² passed the Turing test, marking an artificial intelligence milestone. In this instance, Eugene fooled a third of the human judges it interacted with into believing that that they were messaging a 13-year-old boy from Ukraine, rather than a piece of software.⁶²³ The growth of these

⁶²¹ Gewin, V. (2016). *Data sharing: an open mind on open data*. *Nature*, 529(7584), 117-119.

⁶²² Interview with Eugene Goostman, the Fake Kid Who Passed the Turing Test [Internet]. Time.com. 2017 [cited 5 September 2017]. Available from: <http://time.com/2847900/eugene-goostman-turing-test/>

⁶²³ Howard-Smith F. Humanity Losing its touch [Internet]. Linked In. 2017 [cited 5 September 2017]. Available from: <https://www.linkedin.com/pulse/humanity-losing-its-touch-fabia-howard-smith>

technologies will undoubtedly continue to refine and advance in the next 5, 10 and 20 years. Producing humanly complex AI.

2.2 Is the current level of excitement which surrounds artificial intelligence warranted?

2.2.1 *Origin for the excitement that surrounds AI*

The purpose of designing artificial intelligence software is to develop a system able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, language translation and decision-making. In part, this purpose has led scientists to create software that can perform phenomenological tasks, previously thought to be reserved only for humans. Not only is this a huge technological feat, but is continuing to astound us on a daily basis.

2.2.2 *Is the excitement warranted?*

Google's aforementioned AlphaGo system is the most famous example and a prime credential of why the field of AI is currently exhibiting a period of excitement. AlphaGo was built using what is known as unsupervised learning algorithms. Essentially, this form of learning is akin to a computer finding patterns in data with little or no guidance. Scientists often refer to this type of AI using a black box algorithm - such that, we know the inputs and outputs of the AI, but it is currently unknown how the system arrives at its outcome. This gap in knowledge means, regardless of the probability, there is a possibility that AI could develop uncontrollably. This has caused the formation of organisations with the sole aim of monitoring any developments in the field – e.g. OpenAI. In practice, there is no public examples of artificial intelligent software that are capable of anything more than recognising patterns in games, images or blocks of text.

2.2.3

From an economic standpoint, corporations are looking towards AI for the purposes of automation. Whilst using software to automate is not a new idea, the recent interest found in corporations has spawned because of the introduction of new technologies such as robotic process automation (RPA).⁶²⁴ By 2019, RPA is projected to impact 44% of the total jobs in Australia.⁶²⁵ Recent research by Accenture has shown that RPA can reduce costs by 80% and reduce the time to

⁶²⁴ Lacity, M., Willcocks, L. P., & Craig, A. (2015). Robotic process automation at Telefonica O2.

⁶²⁵ People, Change and Robots [Internet]. 2016 [cited 6 September 2017]. Available from: <https://www.pwc.com.au/pdf/robotic-process-automation-people-change-and-robots.pdf>

perform a task by 80-90%.⁶²⁶ Furthermore, it has been published⁶²⁷ that businesses that successfully apply artificial intelligence (AI) could increase profitability by an average of 38 percent by 2035. These figures indicate the use of AI could be one of the biggest cost saving activities for corporations within the future. Thus there is an understandable excitement from the perspective of industry executives.

3. Industry

3.1 What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

3.1.1 Key sectors that stand to benefit from the adoption of artificial intelligence

Accenture LLP has published research stating that the adoption of AI could lead to an economic boost of US\$14 trillion, affecting 16 industries and 12 economies by 2035.⁶²⁸ The logistics industry provides a prime example of how the rise in automation caused by AI can influence not only an entire industry, but also introduce tremendous societal and economical effects. The UK's Logistics & Post sector is worth approximately £55bn to the economy and comprises 5% of the UK GDP. Currently 1.7m people are employed in this sector by over 63,000 companies.⁶²⁹

3.1.2

Since 2015, the Transport Research Laboratory (partly funded by the UK's Engineering and Physical Sciences Research Council (EPSRC), has focused on understanding the implications of the introduction of autonomous vehicles.⁶³⁰ If the EPSRC deem autonomous vehicles to become safer than human operated vehicles, which aligns with research produced by the Transportation Institute of Virginia Tech and Baidu⁶³¹, a large proportion of jobs within the logistics sector will be at risk if regulations are slow to adapt. Additionally, the lack of regulation

⁶²⁶ Robotic Process Automation | Accenture [Internet]. Accenture.com. 2017 [cited 6 September 2017]. Available from: <https://www.accenture.com/no-en/insight-financial-services-robotic-process-automation>

⁶²⁷ Purdy M, Daugherty P. HOW AI INDUSTRY BOOSTS PROFITS AND INNOVATION [Internet]. Accenture.com. 2017 [cited 6 September 2017]. Available from: https://www.accenture.com/t20170620T055506Z__w_/us-en/_acnmedia/Accenture/next-gen-5/insight-ai-industry-growth/pdf/Accenture-AI-Industry-Growth-Full-Report.pdf?la=en

⁶²⁸ Accenture Report: Artificial Intelligence Has Potential to Increase Corporate Profitability in 16 Industries by an Average of 38 Percent by 2035 | Business Wire [Internet]. Businesswire.sys-con.com. 2017 [cited 6 September 2017]. Available from: <http://businesswire.sys-con.com/node/4109333>

⁶²⁹ Industry Sector Guide – Transport & Logistics August 2017 [Internet]. Ctp.org.uk. 2017 [cited 6 September 2017]. Available from: <https://www.ctp.org.uk/assets/x/53133>

⁶³⁰ TRL to contribute to £11m autonomous vehicle research programme [Internet]. TRL. 2015 [cited 6 September 2017]. Available from: <https://trl.co.uk/news/prev/32021>

⁶³¹ West, D. M. (2016). Moving forward: Self-driving vehicles in China, Europe, Japan, Korea, and the United States.

within the UK concerning the development of self-driving cars by corporations could pose a safety hazard to the general public. The US government have recently proposed a bill to respond to such safety concerns (the bill is outlined in section 6). And as the logistics sector currently faces issues with fuel prices, climate change, increased competition and low growth, corporations will be looking for the next crucial cost cutting exercise.

4. Ethics

4.1 What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

4.1.1 Ethical implications of the development and use of artificial intelligence in regards to Autonomous Vehicles (AVs)

As the presence of AVs grace our roads in the near future, the world changing benefits will be undeniable. But first, there is a very real ethical dilemma we're faced with - the age old philosophical debate that originates from the trolley dilemma. The 'Trolley Experiment'⁶³² was developed to study the morality of action versus inaction. The experiment theorises a runaway trolley racing down a track towards several people. There is a lever next to you which, when pushed, will change the route of the trolley. The alternate direction only has one person on the track. You have two options: 1) don't act, while the trolley kills the five people or 2) act, letting the trolley kill one person in order to save five. This example is often tweaked to exacerbate the problem i.e. five elderly people versus one child? What if the five people were criminals?

4.1.2

AVs will be in similar ethical situations, having to choose between an infant's life running in the road or the passengers' lives, swerving into a road barrier. The result is either harming and maybe even killing the car's passengers or the infant. This dilemma raises the question - are AV's programmed to be deontological or utilitarian? What should it do in an unavoidable accident? The answers to these questions will determine the acceptance and adoption of AVs in today's society. A paper titled 'The social Dilemma of autonomous vehicles'⁶³³ published by Bonnefon from the Toulouse School of Economics tries to answer these questions, it concluded that the public majority showed a moral preference for utilitarian AVs (AVs that are programmed to benefit the majority and minimise the number of casualties). Despite this, participants preferred the self-protective AVs for themselves. It's all well and good discussing this, but it is key to remember that the ultimate goal for AVs is not to get to the point of choosing

⁶³² Liao S, Wiegmann A, Alexander J, Vong G. Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology*. 2012; 25(5):661-671.

⁶³³ Bonnefon J, Shariff A, Rahwan I. The social dilemma of autonomous vehicles. *Science*. 2016; 352(6293):1573-1576.

which group of human are more ethical to sacrifice, yet it is still a question that needs to be factored into the algorithm.

4.1.3 How can any negative safety implementations be resolved

The advantages of self-driving cars are numerous, and therefore can discount any negative safety benefits associated with AVs. The number of lives saved due to human-error related accidents will be significant. AVs are consistent, analytical, unable to get drunk, angry, distracted and do not consume drugs - these are all factors that lead to road fatalities every year. Over 22,000 people were seriously injured in road accidents in the United Kingdom in 2015⁶³⁴, and nearly 1.3 million people worldwide were killed.⁶³⁵ Not only would the number of human casualties be reduced with the introduction of AVs, but increased traffic efficiency⁶³⁶, reduced traffic accidents by up to 90%⁶³⁷ and reduced pollution levels⁶³⁸ will all be possible. There's no doubt that manufacturing ethically minded AVs will be one of the key hurdles artificial intelligence faces today. These type of AI related studies requires an in-depth study into algorithmic morality and showcases a growing need for a national ethics board for AI. Despite this, introducing AV technology too slowly would be unethical due to the overwhelming safety benefits that self-driving cars will bring to our roads.⁶³⁹

4.2.1 In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

4.2.2 Black box definition and transparency concerns

Black box algorithms derive its name from 'black boxes'; something known fundamentally and solely for its inputs and outputs. Users, and even the programmers who create the AI algorithms aren't aware of the internal workings of black box algorithms. Such a lack of transparency must be addressed before implementing artificial intelligence on a large scale. Transparency as a concept is without its criticisms. The Select Committee are advised not to overlook this, questions such as: what specifications need to be met for something to be

⁶³⁴ (<https://www.gov.uk/government/statistics/reported-road-casualties-in-great-britain-main-results-2015>)

⁶³⁵ Road Crash Statistics [Internet]. Asirt.org. 2017 [cited 28 June 2017]. Available from: <http://asirt.org/initiatives/informing-road-users/road-safety-facts/road-crash-statistics>

⁶³⁶ Van Arem B, Van Driel CJ, Visser R. The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, 7:429–436, 2006.

⁶³⁷ Gao P, Hensley R, Zielke A. A roadmap to the future for the auto industry, 2014.

⁶³⁸ Spieser K et al. Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: A case study in Singapore. In Meyer G, Beiker S, editor, *Road Vehicle Automation*, pages 229–245. Springer, 2014.

⁶³⁹ Howard-Smith F. Programmed to Kill [Internet]. Linked In. 2017 [cited 5 September 2017]. Available from: <https://www.linkedin.com/pulse/programmed-kill-algorithmic-morality-self-driving-howard-smith>

considered transparent? And to whom? The organisation that use the AI, the developers who created it or the users whose data it is scraping?

4.2.3 Situations in which a lack of transparency in AI systems is acceptable

In section 2.2 the excitement that surrounds artificial intelligence is warranted on numerous levels and in plethora industries. AI can now perform tasks that would otherwise have required human intelligence, as such, a variety of human jobs can now be replaced by AI systems and to a higher degree of accuracy and efficiency. Situations whereby a lack of transparency in AI systems is acceptable are monotonous and repetitive tasks that don't carry moral weighting, such as the sorting and classification of data. Consider tagging elements of photographs that can later be used in a search function, although a human would be able to complete this task, it will be much speedier and efficient for an AI system to churn through millions of images and correctly tag them. Thus, freeing up humans for more creative and less mundane tasks.

4.2.4 Situations in which a lack of transparency in AI systems should not be permissible

The completion of morally challenging tasks should require a level of transparency available to all people involved. For instance, medical, crime and justice industries are just a few that create hugely important decisions that impact people's lives. However, the fully-fledged human system isn't without its flaws. Medical error is the third leading cause of death in the US, and as many as one in six patients in the British NHS receive incorrect diagnoses.⁶⁴⁰ No wonder statistics like these are raising interest in the artificial intelligence community, as AI systems would be able to dramatically decrease these figures.

The collaboration of both AI and human professionals may be the ultimate answer in these scenarios. For example, an AI could work alongside a doctor when diagnosing a patient. A human doctor can't possibly recall every medical journal ever written, every symptom and corresponding disease, but with the aid of an AI, it can open up diagnoses that weren't considered before. Meanwhile, if the AI's output is something unexpected, it could allow the doctor to investigate that route or overrule the AI's output with their own professional knowledge.

5. The role of the Government

5.1. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

⁶⁴⁰ Hart R. When artificial intelligence botches your medical diagnosis, who's to blame? [Internet]. Quartz. 2017 [cited 6 September 2017]. Available from: <https://qz.com/989137/when-a-robot-ai-doctor-misdiagnoses-you-whos-to-blame/>

5.1.1 The impacts of Artificial Intelligence on UK employment figures should be investigated

Research produced by the US National Bureau of Economic Research has shown that the introduction of a single robot onto an assembly line can result in the loss of up to 5.6 jobs within a local community.⁶⁴¹ Furthermore, every robot introduced within a company per 1,000 workers, reduced the average salary of the local community up to 0.5%. The Select Committee would be wise to replicate the research within the UK, in order to understand the impacts that the rising level of automation is having on the UK's economy and society. Only after researching the current impacts that automation is having on UK unemployment rates, will the Select Committee have grounding for a regulatory response to corporations' use of AI.

6. Learning from others

6.1 What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

6.1.1 American policies for Autonomous Vehicles

In recent months the American government have acknowledged the developments of autonomous vehicles as an important technological milestone and responded with legislation. In June 2017, the US House Energy and Commerce Committee approved a revised bipartisan bill that would speed the deployment of self-driving cars without human controls and bar states from blocking autonomous vehicles.⁶⁴² The purpose of the bill is to allow car manufacturers the ability to develop self-driving cars safely, with significant government oversight. Currently, there is little regulation within the UK for the development of self-driving cars for automobile corporations. Should the Select Committee wish to propose guidelines on the development of safe autonomous vehicles, the UK would have greater control on the advancements within this sector.

6 September 2017

⁶⁴¹ Acemoglu, D., & Restrepo, P. (2017). Robots and Jobs: Evidence from US labor markets.

⁶⁴² Shepardson D. House panel approves legislation to speed deployment of self-driving cars [Internet]. Reuters. 2017 [cited 6 September 2017]. Available from: <https://www.reuters.com/article/us-usa-selfdriving-vehicles/house-panel-approves-legislation-to-speed-deployment-of-self-driving-cars-idUSKBN1AC2K0>

The Human Rights, Big Data and Technology Project – Written evidence (AIC0196)

SUBMISSION TO THE HOUSE OF LORDS SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE BY THE HUMAN RIGHTS, BIG DATA AND TECHNOLOGY PROJECT (HRBDT)

6 September 2017

Executive Summary

- Artificial intelligence already plays a central role in public and private life and this role is only set to increase.
- While artificial intelligence may offer significant benefits to society, particularly if equitably distributed, it also carries significant risk, including to the protection and promotion of human rights.
- Debates on how to ensure that artificial intelligence benefits all in society and contributes to, rather than threatens, human rights are only beginning to take place. Often framed as the need for an ‘ethical approach’ to the development and use of artificial intelligence, no consensus has been reached on what such an ‘ethical approach’ entails.
- HRBDT submits that international human rights standards and norms should sit at the heart of such an approach. Currently, the right to privacy or the prohibition of discrimination in algorithmic decision-making are often treated synonymously with human rights. However, the international human rights framework is much broader than these two rights alone. It encompasses a range of substantive and procedural rights, as well as setting out the obligations of duty bearers (States and corporations) and rights of those affected. It also requires a systematic and integrated approach to the prevention, monitoring and oversight, and accountability and remedies for human rights concerns. It therefore provides a clear, internationally agreed and effective solution to address many of the questions facing the development and use of artificial intelligence.
- In the view of HRBDT, a human rights-based approach should sit at the centre of the development and use of artificial intelligence, enabling a more holistic, consistent, universal, and enforceable approach.

A. Introduction

1. This submission is made by the Human Rights, Big Data and Technology Project (‘HRBDT’), funded by the Economic and Social Research Council and housed at the University of Essex’s Human Rights Centre.⁶⁴³

⁶⁴³ *HRBDT website*. Available at: <www.hrbdt.ac.uk> [last accessed 06.09.17].

2. HRBDT analyses the challenges and opportunities presented by big data and associated technologies from a human rights perspective. It considers both the threats posed to human rights, as well as whether big data and associated technologies can advance the protection and promotion of human rights. Drawing on the expertise of its interdisciplinary team of researchers and partner organisations, HRBDT considers whether fundamental human rights concepts and approaches need to be adapted to meet the rapidly evolving technological landscape. The work brings together States, business enterprises, United Nations officials, practitioners, civil society and academics in the fields of human rights, big data and associated technologies to assess existing regulatory responses and whether reforms are needed in order to maximise effective human rights protection.
3. HRBDT is grateful to the Select Committee on Artificial Intelligence for the opportunity to make this submission. In our submission, we suggest that a human rights-based approach can facilitate a more effective response to both the positive and negative human rights implications arising from the development and deployment of artificial intelligence.
4. For the purpose of this submission, the term ‘artificial intelligence’ has been interpreted broadly to include autonomous or semi-autonomous non-programmed decision making applications and/or systems, which includes an element of machine learning.⁶⁴⁴

B. Potential Opportunities and Risks of Artificial Intelligence for Human Rights

5. Society is only just beginning to understand the impact of artificial intelligence on how we live and interact. Technology is becoming more powerful, multifunctional and central by the day. Smart cities and the Internet of Things are revolutionising our home lives and city planning. Automation and algorithmic decision-making have already reduced human input in both the public and private sectors. The development of autonomous and semi-autonomous systems has the potential to reduce human input even further. These systems are enabled by the amount and type of information that can be collected, amalgamated and stored, and the unprecedented capacity to produce and process large datasets to uncover patterns and correlations that were previously not possible.

⁶⁴⁴ See, e.g., Big Innovation Centre, *What is AI? A Theme Report Based on the 1st Meeting of the All-Party Parliamentary Group on Artificial Intelligence* (2017) pgs. 9 – 13.

6. Developments in artificial intelligence could have a positive impact on the protection and promotion of human rights and offer significant benefits to society, although this is dependent on the benefits being accessible and available equitably within society and beyond the global North.⁶⁴⁵ The sharpest illustration of this is the potential centrality of artificial intelligence, and its capitalisation of big data, to the United Nations Sustainable Development Goals.⁶⁴⁶
7. However, this paradigm shift in technological ability and capacity presents significant risks to human rights. The right to privacy, freedom of expression and association, and equality and non-discrimination are most often cited as at risk from artificial intelligence. However, the risks presented by artificial intelligence threaten the full panoply of rights.⁶⁴⁷ These risks are heightened through hacking and denial of service attacks on critical infrastructure, highlighting the importance of digital security in underpinning the enjoyment of human rights.
8. The risks and the opportunities for the protection and promotion of human rights by artificial intelligence raise the fundamental question of how the opportunities can be maximised while minimising the risks of the digital age. This has generated significant discussion, with possible solutions tending to be characterised as the need for an 'ethical approach' to artificial intelligence. Indeed, in the last few years, literature and policy materials seeking to advance ethical approaches to the development and use of artificial intelligence have burgeoned.⁶⁴⁸

⁶⁴⁵ For more information, please see: V Ng and C Kent, *Human Rights in the Digital Age: The Promises of Big Data and Technology (Part I)* (2016) available at: <www.hrbdt.ac.uk/human-rights-in-the-digital-age-the-promises-of-big-data-and-technology-part-i/> [last accessed 06.09.17].

⁶⁴⁶ See, e.g., *AI for Global Good Summit* (7 – 9 June 2017) available at: <www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx> [last accessed 06.09.17]; Independent Expert Advisory Group on a Data Revolution for Sustainable Development, *A World That Counts: Mobilising the Data Revolution for Sustainable Development* (2014) available at: <www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf> [last accessed: 06.09.17].

⁶⁴⁷ For more information, please see: V Ng and C Kent, *Human Rights in the Digital Age: The Perils of Big Data and Technology (Part II)* (2016) available at: <www.hrbdt.ac.uk/human-rights-in-the-digital-age-the-perils-of-big-data-and-technology-part-ii/> [last accessed 06.09.17].

⁶⁴⁸ See, e.g., IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (Version 1- For Public Discussion)* (2016) available at: <http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf> [last accessed 06.09.17]; *Asilomar AI Principles* (2017) available at: <<https://futureoflife.org/ai-principles/>> [last accessed 06.09.17]; European Parliament Policy Department for Citizens'

9. However, to date, discussions regarding an ethical approach to the development and use of artificial intelligence have been relatively fluid. Some common principles are beginning to emerge.⁶⁴⁹ However, no clear consensus on what an ethical approach entails has yet been reached.
10. In the view of HRBDT, an 'ethical approach' to the development and use of artificial intelligence should not be presented as starting from scratch. Rather, international human rights standards and norms are already well-developed and internationally agreed and can adapt and respond to the challenges of the digital age. They can therefore sit at the centre of the development and use of artificial intelligence, in the form of a human rights-based approach.
11. A human rights-based approach to artificial intelligence would entail 'a process that adheres to both the values which underpin human rights law as well as their substantive content'.⁶⁵⁰ It incorporates key principles,⁶⁵¹ as

Rights and Constitutional Affairs, *European Civil Law Rules in Robotics: Study for the JURI Committee* (2016) available at: [www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU\(2016\)571379_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf) [last accessed 06.09.17]; European Parliament, *Robots and Artificial Intelligence: MEPs Call for EU-Wide Liability Rules* (16 February 2017, European Parliament Press Room) available at: www.europarl.europa.eu/news/en/press-room/20170210IPR61808/robots-and-artificial-intelligence-meps-call-for-eu-wide-liability-rules [last accessed 06.09.17]; House of Commons Science and Technology Committee, *Robotics and Artificial Intelligence: Fifth Report of Session 2016 – 2017* (2016) available at: publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf [last accessed 06.09.17]; Royal Statistical Society, *The Opportunity and Ethics of Big Data: Workshop Report* (2016) available at: www.rss.org.uk/Images/PDF/influencing-change/2016/rss-report-opps-and-ethics-of-big-data-feb-2016.pdf [last accessed 06.09.17]

⁶⁴⁹ Such as responsibility, transparency and human benefit/values. See, e.g., IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (Version 1- For Public Discussion)* (2016) available at: http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf [last accessed 06.09.17]; *Asilomar AI Principles* (2017) available at: <https://futureoflife.org/ai-principles/> [last accessed 06.09.17].

⁶⁵⁰ Northern Ireland Public Services Ombudsman and Northern Ireland Human Rights Commission, *Human Rights Manual* (2016).

⁶⁵¹ Such as participation, accountability, non-discrimination, transparency, human dignity, empowerment and rule of law: Food and Agriculture Organization of the United Nations, *Human Rights Principles: PANTHER* (2006) available at: www.fao.org/righttofood/about-right-to-food/human-right-principles-panther/en/ [last accessed 06.09.17]. See also, Scottish Human Rights Commission, *A Human Rights Based Approach: An Introduction [PANEL Principles]* (undated) available at:

well as substantive and procedural international human rights standards and norms. It focuses on both rights-holders and corresponding obligations of duty-bearers across the cycle of human rights concerns - from prevention, to monitoring and oversight, to accountability and remedies. A human rights-based approach provides a system that can be applied to plans, policies and processes in order to ensure that those most centrally affected are considered and centrally involved.⁶⁵²

C. A Human Rights-Based Approach to the Development and Use of Artificial Intelligence: Application in Theory

12. Existing proposals for an 'ethical approach' to the development and use of artificial intelligence contain some of the key principles and values on which human rights are based, such as human dignity, autonomy and empowerment, and in some cases make reference to human rights, particularly the right to privacy and non-discrimination.⁶⁵³ However, they typically do not incorporate the full framework, take a systematic approach from prevention to remedy or focus on the duty bearers and rights-holders.

13. A human rights-based approach that draws on existing international human rights standards and norms provides enhanced certainty and ensures international perspectives that are based on universal values.⁶⁵⁴

<www.scottishhumanrights.com/media/1409/shrc_hrba_leaflet.pdf> [last accessed 06.09.17].

⁶⁵² For an example of a human rights-based approach to data in the context of the United Nations Sustainable Development Goals, please see: OHCHR, *A Human Rights Based Approach to Data: Leaving No One Behind in the 2030 Development Agenda: Guidance Note to Data Collection and Disaggregation* (2016) available at: <<http://hrbaportal.org/wp-content/files/GuidanceNoteonApproachtoData.pdf>> [last accessed 05.09.17]. For an example of a human rights-based approach to data sharing, please see: T Harris and J Wyndham, 'Data Rights and Responsibilities: A Human Rights Perspective on Data Sharing' (2015) *10(3) Journal of Empirical Research on Human Research Ethics* 334 – 337.

⁶⁵³ See, e.g., IEEE, *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (Version 1- For Public Discussion)* (2016) available at: <http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf> [last accessed 06.09.17]; *Asilomar AI Principles* (2017) available at: <<https://futureoflife.org/ai-principles/>> [last accessed 06.09.17].

⁶⁵⁴ See, e.g., *Vienna Declaration and Programme of Action* [Adopted by the World Conference on Human Rights in Vienna on 25 June 1993] available at: <www.ohchr.org/EN/ProfessionalInterest/Pages/Vienna.aspx> [last accessed on 06.09.17]; Office of the United Nations High Commissioner for Human Rights, *The Core International Human Rights Instruments and their Monitoring Bodies* (undated) available at: <www.ohchr.org/EN/ProfessionalInterest/Pages/CoreInstruments.aspx> [last accessed 06.09.17].

This reduces issues associated with identifying a shared understanding regarding the content of ethical principles and also with the possibility of fragmented and divergent approaches.

14. The existing international human rights framework also unifies the concept of duty-bearers. International human rights law provides a legally binding obligation on States to respect, protect and fulfil human rights. Additionally, under the UN Guiding Principles on Business and Human Rights, businesses have a responsibility to respect human rights.⁶⁵⁵ For example, Internet intermediaries and social media platforms are increasingly involved in online content regulation, which is often automated through artificial intelligence systems. This has potential impacts for freedom of expression.⁶⁵⁶ Such content determinations, which were traditionally in the domain of the State, represent the growing public function of such entities. It is crucially important that such decisions are made in accordance with human rights standards, rather than, for example, 'community standards'. A human rights-based approach offers increased transparency within policy formulations, and 'empowers people and communities to hold those who have a duty to act accountable'.⁶⁵⁷
15. Finally, a human rights-based approach provides a holistic view and enables appreciation of the 'social, political and legal' landscape, thereby '[lifting] sectoral "blinkers" and [facilitating] an integrated response to multifaceted...problems'.⁶⁵⁸ A human rights-based approach extends beyond a compliance mentality, providing a more substantive mechanism by which to identify, prevent and mitigate risk.
16. While these reasons are offered, there remains the need for further research into the nexus between human rights and ethics in the context of the digital age, focusing on potential areas of overlap that may lack clarity and/or produce tensions due to differing approaches.

⁶⁵⁵ Office of the United Nations High Commissioner for Human Rights, *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework* (2011) available at:

<www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf> [last accessed 06.09.17], Section II.

⁶⁵⁶ See, e.g., S Cope et al., 'Industry Efforts to Censor Pro-Terrorism Online Content Pose Risks to Free Speech' (*EFF*, 12 July 2017) available at: <www.eff.org/deeplinks/2017/07/industry-efforts-censor-pro-terrorism-online-content-pose-risks-free-speech> [last accessed 06.09.17].

⁶⁵⁷ Office of the United Nations High Commissioner for Human Rights. *Frequently Asked Questions on a Human Rights-Based Approach to Development Cooperation* (United Nations, New York and Geneva, 2006) available at: www.ohchr.org/Documents/Publications/FAQen.pdf [last accessed on 05.09.17] pg. 17.

⁶⁵⁸ *Ibid.* pg. 17.

D. A Human Rights-Based Approach to the Development and Use of Artificial Intelligence: Application in Practice

17. A human rights-based approach cuts across a wide range of considerations relevant to the development and use of artificial intelligence, including prevention and protection, due process and access to information, responsibility and accountability, access to justice, oversight and remedies.⁶⁵⁹ Artificial intelligence programming, processes, policies and planning should be guided at all stages by 'human rights standards as reflected in the international treaties, as well as principles such as participation, non-discrimination and accountability'.⁶⁶⁰
18. The following are 'necessary, specific and unique' to a human rights-based approach:
- (a) 'Assessment and analysis in order to identify the human rights claims of rights-holders and the corresponding human rights obligations of duty-bearers, as well as the immediate, underlying, and structural causes of the non-realisation of rights;
 - (b) [Assessment of the] capacity of rights-holders to claim their rights, and of duty-bearers to fulfil their obligations, [followed by the development of] strategies to build these capacities;
 - (c) [Monitoring and evaluating] both outcomes and processes guided by human rights standards and principles;
 - (d) [Programming, processes, policies and planning are] informed by the recommendations of international human rights bodies and mechanisms.⁶⁶¹

⁶⁵⁹ See, in the context of data protection: UNHCR, *Policy on the Protection of Personal Data of Persons of Concern to UNHCR* (May 2015) available at: www.refworld.org/docid/55643c1d4.html [last accessed on 05.09.17]. This policy 'represents the first attempt by a UN agency to adopt a comprehensive, principled and universal approach to data protection', including a 'ground breaking' section on individual rights that represented an 'important contribution to make international organisations more accountable to respect individual rights' (A Beck and C Kuner, 'Data Protection in International Organizations and the New UNHCR Data Protection Policy: Light at the End of the Tunnel?' (31 August 2015, *EJIL: Talk!*) available at: www.ejiltalk.org/data-protection-in-international-organizations-and-the-new-unhcr-data-protection-policy-light-at-the-end-of-the-tunnel/ [last accessed 05.09.17].

⁶⁶⁰ Office of the United Nations High Commissioner for Human Rights. *Frequently Asked Questions on a Human Rights-Based Approach to Development Cooperation* (United Nations, New York and Geneva, 2006) available at: www.ohchr.org/Documents/Publications/FAQen.pdf [last accessed on 05.09.17] pg. 23.

⁶⁶¹ *The Human Rights-Based Approach to Development Cooperation: Towards a Common Understanding Among the United Nations Agencies* (Second Inter-Agency Workshop, Stamford,

19. The application of a human rights-based approach to accountability in the context of artificial intelligence-based algorithms provides illustration of how this approach would apply in the realm of artificial intelligence. As has been well documented, algorithmic decision making has the potential to 'bake-in' and potentially exacerbate existing inequalities and discrimination.⁶⁶² For example, Kroll argues that: first, 'algorithms that include some types of machine learning can lead to discriminatory results if the algorithms are trained on historical examples that reflect past prejudice or implicit bias, or on data that offer a statistically distorted picture of groups comprising the overall population';⁶⁶³ second, 'machine learning models can build in discrimination through choices in how models are constructed' (input data, proxies etc);⁶⁶⁴ and third, 'there is the problem of "masking": intentional discrimination disguised as one of the above mentioned forms of unintentional discrimination'.⁶⁶⁵

20. In this context, a human rights-based approach requires using international human rights standards and norms as a means for identifying and defining elements within the algorithm life-cycle that give rise to human rights concerns, establishing which entity/entities impact(s) upon rights, addressing questions of responsibility, and identifying how human rights concerns can be addressed. In the context of inequality and discrimination, HRBDT has previously proposed that 'the design and testing of algorithms should be approached with a view to prioritising the

USA, May 2003) available in Annex II at:

www.ohchr.org/Documents/Publications/FAQen.pdf [last accessed on 05.09.17] [3].

⁶⁶² See, e.g., Council of Europe, *Study on the Human Rights Dimensions of Algorithms: Second Draft* (20 February 2017) available at: rm.coe.int/16806fe644; L Rainie and J Anderson, 'Code-Dependent: Pros and Cons of the Algorithm Age' (*Pew Research Center*, 8 February 2017) available at: assets.pewresearch.org/wp-content/uploads/sites/14/2017/02/08181534/PI_2017.02.08_Algorithms_FINAL.pdf;

J Angwin et al., 'Machine Bias' (*ProPublica*, 23 May 2016) available at:

www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing;

The White House, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (May 2016) available at:

obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf;

C O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Publishing, 2016);

Center for Democracy and Technology, *Digital Decisions* (undated) available at: cdt.org/issue/privacy-data/digital-decisions/;

F Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015).

⁶⁶³ J Kroll et al., 'Accountable Algorithms' (2017) 165 *University of Pennsylvania Law Review* 633, 680.

⁶⁶⁴ *Ibid.*, 681.

⁶⁶⁵ *Ibid.*, 682.

prevention or at least minimisation of discrimination in outcomes. This could yield different results than using abstract proxy metrics for distinct groups to assess discrimination theoretically'.⁶⁶⁶

21. A human rights-based approach also offers a framework of obligations that maps across the algorithmic life-cycle from prevention to remedies and ex post facto redress, building on a shared framework of State responsibility and business and human rights standards, cognisant of the multi-stakeholder ecosystem in this context. In this regard, as previously highlighted by HRBDT, '[g]reater clarity is needed on the interaction between manual human evaluation and automated algorithmic systems, and how safeguards can be implemented'.⁶⁶⁷

22. A human rights-based approach to the accountability of algorithmic decision-making will include tools such as initial and ongoing human rights impact assessments to test and review the impact of algorithmic decision-making on human rights. Human rights impact assessments are 'instruments for examining policies, legislation, programmes and projects prior to their adoption to identify and measure their impact on human rights... They are designed to identify the intended and unintended impact on the enjoyment of human rights, and the State's ability to protect and fulfil them. As such, they are a planning tool to prevent human rights violations by assessing the formal or apparent compatibility of laws, policies, budgets and other measures with human rights obligations, as well as the likely impact in practice, thus creating the opportunity for reconsideration, revision or adjustment prior to adoption. Prior consultation with relevant stakeholders can assist in identifying possible impacts on human rights'.⁶⁶⁸ Undertaking such efforts can complement the artificial intelligence design and deployment process, instead of focusing solely on post-facto accountability, which expands the opportunity for accountability beyond current outcome-based approaches.⁶⁶⁹

⁶⁶⁶ HRBDT, *Written Evidence Submitted to the Science and Technology Committee Inquiry on Algorithms in Decision-Making* (26 April 2017) available at: <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/algorithms-in-decisionmaking/written/69117.pdf> [last accessed 06.09.17].

⁶⁶⁷ *Ibid.*

⁶⁶⁸ UN General Assembly, *The Role of Prevention in the Promotion and Protection of Human Rights: Report of the Office of the United Nations High Commissioner for Human Rights* (16 July 2015) UN Doc A/HRC/30/20 [31].

⁶⁶⁹ J Kroll et al., 'Accountable Algorithms' (2017) 165 *University of Pennsylvania Law Review* 633, pgs. 682 – 692; C Dwork et al., 'Fairness Through Awareness' (2012) *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* 214 - 226.

23. There are numerous questions requiring consideration in relation to the practical application of a human rights-based approach to the development and use of artificial intelligence. For example, within the prevention phase, part of the algorithmic impact assessment will be to determine whether or not it is appropriate to deploy an artificial intelligence-based algorithm to address the specific issue at hand. Should such an algorithm ever be used to make a decision affecting, for example, an individual's right to liberty, given that such technology is typically based on population, not individual, level trends and on correlation, not causation? Answering this and related questions will require the establishment of criteria for deciding whether or not it is appropriate for algorithms to be used when they have serious human rights impacts. Such questions, which are all the more important in light of increasing automation, will require a multi-stakeholder approach in order to ensure that both the challenges facing duty-bearers in the exercise of their obligations are adequately addressed and overcome, and that rights-holders have the capacity, and are empowered to, claim their rights.

6 September 2017

Dr Catrin Fflûr Huws – Written evidence (AIC0008)

An area of artificial intelligence in which I am particularly interested is its use within law, and therefore my comments focus on question 2, 3, 4, 6 and 8. Although much work has been done in terms of using artificial intelligence and machine learning to identify relevant case precedents, my recent work on artificial intelligence and law (Huws, C.F. and Finnis, JC (2017) *On Computable Numbers with an Application to the Alan Turing Problem 25(2) Artificial Intelligence and Law 181*, has identified some of the limitations of applying artificial intelligence in a legal context. These are broadly summarised as follows:

- a. Irrespective of whether specific legislation criminalises specific behaviour, the question of whether that conduct is in fact pursued by the police and prosecuted depends on a number of factors outside the legal system – this may depend on social mores and the attitudes of individuals in influential situations, other pressures on the authorities and the conduct and the identity of the individual. Therefore, although a machine may be programmed to consider particular variables, the programmer is not able to predict which external influences will affect the operation of the law.
- b. In many legal situations, an issue is disputed because there are very fine nuances of behaviour. In the law of tort for example, a decision regarding whether a defendant's conduct was an act of negligence will sometimes depend on very fine gradations of behaviour as questions of what is, for example, reasonable may vary according to macro level conceptions of what constitutes appropriate behaviour, as well as micro level conceptions of what was reasonable for that time, place and context. The ability of a machine to identify appropriate precedents is therefore likely to be something of a blunt instrument.
- c. There are instances where the courts have not followed the relevant precedents. This may arise in situations where there is a realisation that the law has not kept pace with social attitudes. The case of *R v R (Rape: Marital Exception)* [1991] 1 All ER 755 may be an example here, where the machine's understanding of the law would be to decide the case in a manner that is consistent with the relevant precedents. However, the human court had the ability to act in a way that was unpredictable from the case precedents. Furthermore, there are situations where, despite all evidence indicating a guilty verdict, a jury has declined to find the defendant guilty.
- d. The law also operates with reference to different people's understanding of what words mean. Accordingly, the expertise of discourse analysts indicates that that which the speaker intends to communicate may be different from that which the listener understands. Therefore, although a

machine could learn how specific words and concepts have been interpreted in earlier cases, a machine's ability to evaluate whether that is the only possible meaning, having regard to the reader or listener's contextual inputs may be more problematic.

- e. The reliance on artificial intelligence and machine learning may also lead to a lack of innovation in legal argument. There may be situation where characterising a legal problem in one way, may overlook the scope for a dispute to be resolved in another way. For example, the development of the law of torts in the 1930s arose because of a judicial willingness to explore a more general concept of a duty of care, rather than confining negligence to specific duty situations. Similarly, the development of the Quistclose trust occurred because legal expertise was able to characterise the problem as a problem of the law of trusts as opposed to a dispute that was confined to the principles of the law of contract. In situations of these types, the law did act in a manner that could have been programmed.
- f. There is also a risk that an undue reliance on artificial intelligence may limit judicial and legal creativity, in that a problem may only be characterised in a manner that is readable by the machine, and that the possible solutions that a machine may suggest are perceived to be the only available solutions to a problem, without there being sufficient scope to consider what alternatives may be available. Computing technology tends to operate on binary classifications, and many of the problematic areas in law arise because a person fits into several categories or no categories. Artificial intelligence may therefore be useful in terms of characterising the type of expertise that might be sought and for identifying possible precedents. It may also be useful in deciding on liability and punishment in strict liability offences.
- g. Often legal terminology has several different meanings and therefore in addition to the scope for artificial intelligence to suggest solutions, and even decide on an outcome, there is also a need to be able to interpret that information, and there is a risk that artificial intelligence may suggest misguided solutions.
- h. Many cases were initially decided on the basis of specific fact situation. The case of *Lloyds Bank v Rosset* is an interesting example [1990] UKHL 14 that many overlook is the fact that the Rossets' house had been put into the sole name of Mr Rosset in order to prevent Mrs Rosset from acquiring a share of the property. The context specific nature of law means that this case may not have become as significant a precedent if this factor had not been present and emphasised to the courts.
- i. Artificial intelligence replicates the assumptions and the perceptions of the programmer in terms of classifying the problem. Therefore, although one

significant advantage of artificial intelligence is said to be the elimination of prejudice, in that the machine will not 'see' the defendant and make assumptions and judgments based on social characteristics, the machine will also be unable to decide the unconscious biases of the programmer – it will learn what is it programmed to learn, and therefore may, for example, exclude the older precedent in favour of the newer precedent, or be unaware of what aspects of the cultural and social context may be relevant.

11 August 2017

IBM – Written evidence (AIC0160)

1. IBM has been researching, developing and investing in AI technology for more than 50 years. The public became aware of a major advance in 2011, when IBM Watson won the historic Jeopardy! exhibition on US television, widely seen at that time to surpass particularly difficult AI challenges, such as natural speech recognition. Since that time, the company has advanced and scaled the Watson platform, and applied it to various industries, including healthcare, finance, commerce, education, security, and the Internet of Things. We are deeply committed to this technology, and believe strongly in its potential to benefit society, as well as transform our personal and professional lives.
2. To this end, we have engaged thousands of scientists and engineers from IBM Research and Development, and partnered with our clients, academics, external experts, and even our competitors to explore all topics around AI. We are leveraging our understanding of real world business problems to develop AI systems which address the challenges of a wide range of industry sectors. And we have developed a unique point-of-view, informed by decades of research and commercial application of AI.

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10, 20 years? What factors, technical or societal, will accelerate or hinder this development?

3. The field of Artificial Intelligence (AI) comprises capabilities such as machine learning, reasoning and decision technologies; language, speech and vision technologies; human interface technologies; distributed and high-performance computing; and new computing architectures and devices. When purposefully integrated, these capabilities are designed to solve a wide range of practical problems, boost productivity, and foster new discoveries across many industries. AI systems are already part of our daily lives, answering questions and making recommendations for products and services. More are on the way to help people live and work “smarter” in a world where big data is the new natural resource.
4. AI is making rapid progress with an unprecedented range of applications by applying deep learning techniques to acquire skills automatically from data and from practice. In particular, this has enabled computer performance on perception and skilled action tasks (on which progress to date has been slow and incremental) to rapidly approach human performance levels (speech recognition, object recognition, scene labelling, video game playing). This progress was triggered by the ready availability of large training sets (data) and of supercomputing capacity, but more recently has been accelerated by algorithmic and scientific advances in neural networks, reinforcement learning, and related fields.

5. These advances have resulted in significant achievements in AI capabilities on perceptual tasks related to speech, vision, and audio understanding. Deep learning has essentially helped automate what was previously a human expert process. It is highly likely that deep learning will continue to advance capabilities on perceptual tasks over the next 5, 10, and 20 years. It is also reasonable to expect competent conversational systems in limited domains, and the first highly mobile robots capable of precise manipulations, within five years. We should expect computer and robotic systems that can perform a wide variety of the functions required to run a business, factory or household within ten years. And we may expect very broad AI assistance for complex tasks like management, healthcare, or scientific research within twenty years.
6. The technical factors that will accelerate progress in the next few years come from the continued development of data sets for training, further advances in computing, and sustained innovation on neural network design and deep learning algorithm development built on increasingly available open software/hardware deep learning frameworks.
7. While technical progress will continue to be made on the ability of computers to achieve high accuracy on tasks such as object recognition and tracking, face recognition, language translation, and speech transcription, there are still a number of challenges ahead of us, such as the ability to explain an AI system gave a particular answer (explainability), learning from limited data, combining of knowledge, semantic reasoning and perception, and integration of other important but elusive abilities like common sense with perceptual understanding. Nor is it appropriate to set unreasonable expectations on what technologies like deep learning can do, since challenges like explainability may limit the use of deep learning in practice for important applications, such as medical image diagnosis and autonomous driving.

Is the current level of excitement which surrounds artificial intelligence warranted?

8. We believe that the current level of excitement around AI is warranted. Firstly, the technical community now has access to vastly more computing power and data than was available for previous AI hype cycles. Unprecedented amounts of data and computing infrastructure form the backbone of the training and testing needed for AI to progress. For example, IBM recently reported ⁶⁷⁰ new distributed deep learning software which achieved record scaling efficiency on the recognized Caffe deep learning framework. Other examples of large, cloud-based machine learning computing infrastructures include Google's Tensorflow Processing Units

⁶⁷⁰ https://arxiv.org/abs/1708.02188?cm_mc_uid=&cm_mc_sid_50200000=1474979346

(TPU)⁶⁷¹ Cloud access to these resources means that many more technologists have access to this backbone than ever before. Examples of the unprecedented data include a 100 million Flickr images made available by Yahoo⁶⁷², the 80 million labeled tiny images dataset⁶⁷³, and other sources of faces, fingerprints, and medical images, just to name a few.

9. Secondly, breakthroughs in the training of deep neural networks is enabling tasks to be accomplished through machine learning not previously achieved with AI. Examples include surpassing human error rates for certain perception tasks such as image recognition and speech recognition word error rate, and surpassing human game playing abilities in Jeopardy!, Atari, and Go. In the research community, generative models, previously requiring human curation, are now being estimated via an adversarial process referred to as Generative Adversarial Networks⁶⁷⁴.
10. Thirdly, due to these breakthroughs, we are also seeing unprecedented levels of investment in AI techniques. Technology companies like IBM are investing billions in products and research, and investors are pumping even more into a variety of companies with possible AI plays, such as Nvidia, Salesforce, Splunk, Netflix, PayPal, and many others. Research firm Tractica forecasts that worldwide artificial intelligence revenue will reach \$59.8 billion per annum by 2025, up from \$1.4 billion in 2016. IDC, another market research firm, forecasts that AI revenue will grow from \$8 billion in 2016 to more than \$47 billion in 2020, with about half that total being software-related. Funding of AI startups jumped to \$1.73 billion in Q1 2017, up from \$939 million a year earlier, according to CB Insights.
11. These significant investments are enabling not only development and in-market experimentation, but also long-term research needed for continued AI breakthroughs for the foreseeable future.

How can the general public best be prepared for more widespread use of artificial intelligence?

12. Artificial Intelligence is already part of our daily lives, answering questions and giving recommendations, whether this be in internet search engines, GPS mapping systems, anti-virus and malware avoidance or medical devices. There is an acceptance of AI based on a certain level of familiarity and trust in its use. An increased level of attention has come as a result of driverless cars and other eye-catching initiatives. We believe it is important to raise awareness about AI applications with clear societal benefits such as

⁶⁷¹ <https://cloud.google.com/blog/big-data/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu>

⁶⁷² <https://yahooresearch.tumblr.com/post/89783581601/one-hundred-million-creative-commons-flickr-images>

⁶⁷³ <http://groups.csail.mit.edu/vision/TinyImages/>

⁶⁷⁴ https://en.wikipedia.org/wiki/Generative_adversarial_networks

healthcare (IBM's Watson for Oncology being an example) cyber security (such as IBM's Watson for Cyber Security) and education.

13. Data privacy is a crucial issue in everyday AI applications in the UK and European context – for example, chat-bots and automated scoring systems for credit enquiries from the general public. There is a need to address fairness in such systems – in order to prepare the public for the likely increased use, we need to be convinced that these systems are fair and transparent. It is worth noting that AI systems have the capability of being less discriminatory than human beings when they are properly developed. Issues such as the circumstances under which data sets are used or provided to third parties need to be addressed. We go into more detail on both these questions later in our response.
14. Societal barriers to adoption may arise from a public perception that the benefits of AI for the population at large are outweighed by the risks (e.g. increased inequality, automated decision making, privacy violation, corporate and civil control and exploitation). We believe that it is important to take these seriously and we have therefore developed a set of AI Principles⁶⁷⁵ that serve to counter these risks. Ensuring that the benefits of AI are generally and equitably distributed across human society will help in mitigation.

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

15. For decades, we have been stockpiling digital information. We have digitised the history of the world's literature and its medical journals. We track and store the movements of vehicles, trains, planes and mobile phones. And we are privy to the real-time sentiments of billions of people through social media. It is not unreasonable to expect that this rapidly growing body of digital information could deliver significant progress in defeating cancer, reversing climate change, or managing the complexity of the global economy. We believe that many of the ambiguities and inefficiencies of the critical systems that facilitate life on this planet can be eliminated. And we believe that AI systems are the tools that will help us accomplish these ambitious goals.
16. A major bottleneck in developing and validating AI systems is public access to sufficiently large, openly curated, public training data sets. Machine learning requires large, unbiased data sets to train accurate models. Deep learning is advancing speech transcription, language translation, image captioning, and question and answering capabilities. Each new AI advance, e.g., video comprehension, requires the creation of new data sets. Deep domain tasks, such as cancer radiology, or insurance adjustment, require specialised and

⁶⁷⁵ IBM principles for transparency and trust in the cognitive area:
<https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/>

often hard-to-get datasets. Incentives and mechanisms must be created for greater sharing of both input datasets and trained models.

In the Ethics section we include views on dealing the employment and skills implications of AI.

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

17. Industry Examples

- For **healthcare**, AI systems can advance precision medicine by ingesting patient's electronic medical history and relevant medical literature, performing cohort analysis, identifying micro-segments of similar patients, evaluating standard-of-care practices and available treatment options, ranking by relevance, risk and preference, and ultimately recommending the most effective treatments for their patients.
- For **social services**, AI systems can provide timely and relevant answers to citizens in need, assist citizens with insurance, tax, and social programmes, predict the needs of individuals and population groups, and develop plans for efficient deployment of resources.
- For **education**, AI systems can assist teachers in developing personalised educational programmes for individuals or groups of students, assist students using a range of learning styles and methods, and develop effective early education, primary, secondary, and higher education programmes.
- For **financial services**, AI systems can expand financial inclusion by qualifying applicants, assist in providing the best insurance coverage at the right cost, ensure compliance with regulation and reduce fraud and waste in tax and other financial programmes.
- For **transportation**, AI systems can improve the efficiency of public transportation systems, support public vehicles with driver assistance using semi-automated features, manage incidents, optimise the use of fuel and support maintenance of infrastructure and rolling stock.
- For **public safety**, AI systems can support safety personnel with anomaly detection using machine vision, build predictive models for crime, and help investigators find associations in massive amounts of information.
- For the **environment**, AI systems can understand complex relationships and help construct environmental models for accurate prediction and management of pollutants and carbon footprints.
- For **infrastructure**, AI systems can assist with prediction of demand, supply, and use of transport.
- For **manufacturing**, AI systems can dramatically improve production efficiency, reduce waste and increase safety in hazardous environments.

How can the data-based monopolies of some large corporations, and the winner takes all economies associated with them be addressed? How

can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

18. Discussion about what constitutes dominance in data driven economies has focused hitherto on social networks, publicly available search engines, internet based communication and e-commerce web sites – these areas have been subject to competition authority interest in Europe in recent years and there are several well known competition cases currently in progress.

19. It is important, in our view that the custodianship of data and the right to use it is made transparent so data subjects and data rights holders can make informed choices about what can be done with their data. AI services which are provided “for free” in return for user data should be open to scrutiny.

20. The major question with regard to the public good is how personal data is used in AI. Strong data protection safeguards and adequate security mechanisms such as those required by the GDPR are important to strengthen an individual’s privacy rights.

21. Any concerns about dominance can be addressed through competition law on an ex-post basis rather than a-priori regulation of AI. Competition authorities are well equipped to deal with data dominance issues and can monitor the behaviour of companies that amass large amounts of data for commercial exploitation.

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

22. Thanks to significantly improved perception and reasoning capabilities, AI has recently become pervasive in our professional and personal life, providing crucial support for human decision making in diverse areas such as healthcare, social services, education, financial services, transportation, public safety, environment and infrastructure.

23. In order to be fully accepted into society, AI systems need to have significant social capabilities, because their presence in our lives has a profound impact on our emotions and on our decision making capabilities. To achieve this, AI systems also need to understand how to learn and comply with specific behavioral principles to align with human values. AI needs to be equipped with the capability to model and behave according to ethical principles, social norms, professional codes, and moral values suitable for a specific task, context, and culture. Value alignment should become a fundamental topic of research and a requirement for policies and compliance regulations.

24. Moreover, to fully reap the societal benefits of AI, we will first need to trust it. That trust will be earned through experience, and will also require a system of best practice that can guide the safe and ethical development of AI, which

- should also include algorithmic accountability, compliance with existing legislation and policy, and protection of privacy and personal information.
25. Data issues related to privacy, ownership, diversity, and bias are also crucial in AI, since most successful AI systems include a learning module that is heavily trained on available data. AI's decision making or decision support capabilities are as good as the data used to train it. Therefore, it is essential that the data sets used are not biased, otherwise that bias will be transferred into the AI decision process and result. A responsible and ethical development of AI has to include a comprehensive analysis of diversity and bias in data in order to mitigate these possible deficiencies in decision making. Humans have cognitive biases too, and an unbiased AI can also help humans to avoid or mitigate them.
 26. The impact of AI on jobs should also be considered with care. History suggests that powerful technologies like AI result in higher productivity, higher earnings, and overall job growth. We believe that new companies, new jobs, and entirely new markets will be built on the shoulders of this technology. And we believe that AI systems will improve access to critical services for underserved populations. Overall, we anticipate widespread improvements in quality of life. However, AI producers and policy makers have the responsibility to guide the transition and transformation of jobs by helping with reskilling and education, so that as many people as possible can take advantage of the AI revolution.
 27. In order to address all the above issues, IBM has recently published the first principles for transparency and trust in the cognitive area⁶⁷⁶, which are based on a clear purpose for intelligence augmentation rather than replacement, complete transparency on the use of data in building AI, and full commitment to supporting students, workers and citizens acquire the skills and knowledge to engage safely, securely and effectively in a relationship with AI. Moreover, IBM has also published a white paper on trusting AI systems⁶⁷⁷.

In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

28. To achieve the best synergy between human and machine intelligence, we need to build trust in AI systems. Trust is built upon accountability. As such, the algorithms that underpin AI systems need to be as transparent, or at least as interpretable as possible. In other words, they need to be able to explain their behaviour in terms that humans can understand — from how they interpreted their input to why they recommended a particular output.

⁶⁷⁶ IBM principles for transparency and trust in the cognitive area:
<https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/>⁶⁷⁶

⁶⁷⁷ IBM white paper on "Learning to trust artificial intelligence systems":
<http://research.ibm.com/cognitive-computing/ostp/rfi-response.shtml>

29. To do this, we recommend all AI systems should include “explanation-based collateral systems”⁶⁷⁸. The provided explanations should be meaningful to the targeted users. For example, in AI decision support systems whose aim is to help doctors identify the best therapy for a patient, such AI systems need to provide useful explanations to doctors, patients, nurses, relatives, etc. More generally, existing AI systems support many advanced analytical applications for industries like healthcare, financial services and law. In these scenarios, data-centric compliance monitoring and auditing systems can visually explain various decision paths and their associated risks, complete with the reasoning and motivations behind the recommendation. And the parameters for these solutions are defined by existing regulatory requirements specific to that industry.
30. Explanations are definitely needed where laws and regulations require them (such as the GDPR in Europe⁶⁷⁹). However, even when regulations do not require it, we believe that they should be provided to achieve the best collaboration environment for humans and AI, and to create the correct level of trust between them. If explanations are not available, the main risk is that such systems will not be trusted and thus will not be used. We believe that trust is a precursor to adoption, and that adoption is the only path to business success and societal benefits.
31. Since IBM believes all AI systems should always include explanation-based collateral systems, we go beyond transparency about data and algorithms. For cognitive systems to fulfill their world-changing potential, it is vital that people have confidence in their recommendations, judgments and uses. Therefore, IBM is committed to make clear when and for what purposes AI is being applied in the cognitive solutions we develop and deploy. Moreover, we will also clarify the major sources of data and expertise that inform the insights of cognitive solutions, as well as the methods used to train those systems and solutions.

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

32. Responsibility must be the foundation for AI policymaking. Inclusive dialogues can explore relevant topics, going beyond the headlines and hype, promoting deeper understanding and a new skills focus. Every transformative tool that people have created – from the steam engine to the microprocessor – augment human capabilities and enable people to dream bigger and do more. People with these tools will solve whole new classes of big data problems. Our

⁶⁷⁸ Explanation based collateral systems provide guidance as to how decisions were made thus providing explainability and assisting when discrimination or bias in the system needs to be addressed see

http://www.research.ibm.com/software/IBMResearch/multimedia/AIEthics_Whitepaper.pdf

⁶⁷⁹ EU GDPR regulation, which comes into effect in May 2018, and will be mirrored in UK Data Protection law, calls for the right of explanation in AI systems (Recital 71)

responsibility as members of the global community is to ensure, to the best of our ability, that AI is developed the right way and for the right reasons.

33. AI should not be regulated per se. We are very early in the growth curve of AI and the Government should promote policy measures which enable and encourage the adoption of AI. A rush to further regulation can have the effect of chilling innovation and missing out on the societal and economic benefits that AI can bring. Data Privacy regulation, such as GDPR and network and information security regulation, such as the European NIS Directive, as well as product liability laws and consumer protection already provide a legal framework.
34. AI represents a significant economic opportunity for the UK. Policy measures which can help encourage the adoption of AI include: The availability of skills; Incentives to innovate in the UK (financial, legal, and technical); Research and development collaborations. The UK Government could also encourage adoption of AI innovations across the public sector.

What lessons can be learnt from other countries or international organisations in their policy approach to artificial intelligence?

35. From working closely with international organisations such as the European Commission, the World Economic Forum, and the OECD, we believe that a policy approach to AI should be holistic, covering technology, economic, environmental and social issues. A mature, responsible approach to AI should take into account all aspects we have outlined above – the need to address ethical issues, avoiding discrimination through algorithmic transparency, addressing societal needs and dealing with the skills question as well as workplace transformation.
36. The second insight from dealing with international organisations and governments is that the question of ethics and AI cannot be dealt with by one government alone – there is a role for a supra governmental approach, with international organisations working closely with industry to develop high level approaches or codes of practice. In this area, IBM is a founding partner of the Partnership on AI⁶⁸⁰, a multi-stakeholders initiative where both corporate and not-for-profit organisations intend to study and formulate best practices on AI technologies, to advance the public's understanding of AI, and to provide an open platform for discussion and engagement about AI and its influences on people and society.

6 September 2017

⁶⁸⁰ <https://www.partnershiponai.org/>

IEEE European Public Policy Initiative Working Group on ICT – Written evidence (AIC0106)

Author: Petia Georgieva

*University of Aveiro, Portugal
DETI/IEETA - Machine Learning & Intelligent Robotics Lab*

IEEE Senior Member

*On behalf of IEEE European Public Policy Initiative (EPPI) - Working Group on ICT
https://www.ieee.org/about/ieee_europe/europe_ict.html*

1. What is the current state of artificial intelligence and what factors have contributed to this?

The field of Artificial Intelligence (AI) research was formally established at a conference at Dartmouth College in 1956. However, still in 1940, Alan Turing's theory of computation suggested that digital computers could simulate formal reasoning (Turing test). The goal was to investigate ways in which machines could mimic cognitive human functions, such as learning and problem solving. Since then, the field of AI went through ups and downs (such as in the 1960's as well as the mid 1970's) due to over-expectations, limitations in knowledge acquisition and computational resources. Nevertheless, there have been also significant advances and from academic areas of study, nowadays, AI is embedded in mainstream technologies such as robot motion planning and navigation, computer vision (i.e. object recognition), natural language processing and speech recognition, data processing, and knowledge representation and reasoning.

How is it likely to develop over the next 5, 10 and 20 years?

A substantial increase is expected in future applications of AI, including autonomous vehicles (such as drones and self-driving cars), medical diagnosis, treatment and physical assistance for the elderly (e.g. intelligent robotics platforms), smart cities, to mention a few.

However, there is a long way to go before having a real AI that would mimic human reasoning. General AI is still decades away. Currently we develop "narrow AI" systems that perform individual specialized tasks in well-defined domains. These "narrow AI" technologies work and will continue to work alongside humans to extend, augment, enhance human capabilities.

What factors, technical or societal, will accelerate or hinder this development?

If society approaches these technologies primarily with fear and suspicion, missteps that slow AI's development or drive it underground will be the result, impeding important work on ensuring the safety and reliability of AI technologies. On the other hand, if citizens are informed about the positive benefits of AI, while being educated in terms of skills and jobs, the technologies emerging from the field could profoundly transform society for the better in the coming decades⁶⁸¹.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

AI technologies created new challenges for the economy and the society. A common concern about the development of AI is the potential threat it could pose to humankind. The opinion of experts within the AI field is mixed; however, development of militarized artificial intelligence is a commonly shared concern. The United Nations (UN) initiative on banning autonomous weapons⁶⁸² was followed by the open letter of leading AI and Robotics researchers⁶⁸³.

Major fear regarding the AI technology is that it will still jobs, however the humanity has already undergone similar fears when the steam engine was invented, when the electricity was discovered, when the automobile substituted the horse, when the machines largely substituted manual works in agriculture, not that long ago. All these technologies were successful in transforming the society for good, because people found a way to adapt and take advantage of them. Therefore, in order to overcome natural prejudices regarding any new, still not well developed, technology (and AI in particular) and to minimize the inherent risks, we need to learn to dominate it.

Nowadays, AI is the new technology that challenges the society and similarly, as for example with the car invention, we need qualified workforce to develop AI products, to build suitable AI infrastructures, to regulate AI and at the same time we need trained AI customers.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

The impact of AI on the workforce has to be a major concern of the policymakers. The educational agenda needs to be refocused in order to equip the workforce with the necessary digital skills to compete on the free market. AI

⁶⁸¹ <https://ai100.stanford.edu/>

⁶⁸² <https://www.un.org/disarmament/geneva/ccw/background-on-lethal-autonomous-weapons-systems/>

⁶⁸³ <https://futureoflife.org/open-letter-autonomous-weapons/>

coding and understanding of how intelligent systems work need to be trained as new literacy skills for any human. New jobs are required interfacing the AI systems and their end users.

With the emergence of new AI-based industries and, in general, the digital knowledge-based economy, the proportion of the labor force requiring some form of education or training beyond high school will increase significantly. Governments, business, and educational institutions need to share the responsibility for investment in education and training in order to increase the skilled workforce. Pragmatic educative programs need to be designed targeting different generations and different levels of education (not only in the University environment), covering technical and societal (ethical) aspects of AI. Major challenges for policy makers to consider include how training should be organized (e.g. based on projects and learning through examples), by whom (which governmental agency), with what money, who will pay people while they are educated / reskilled.

In the midst of budget deficits and high unemployment, it may be difficult for some to accept the fact that there is a very different and problematic future looming on the horizon. If we do not adopt proactive policies now, we will face a future with large numbers of unskilled workers looking for jobs that require skills they do not possess (people without jobs), and a large number of jobs that will go unfilled (jobs without people).

Providing a highly qualified workforce in AI and providing employment alternatives for those who lose their jobs due to AI are equally urgent labor issues that need to be addressed by the governments. It may be useful to further explore the potential efficacy of the universal basic income experiments recently started in Italy and Finland.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

AI is a technology for creating new consumer products and business applications that are supposed to bring more benefits for the society. AI value and utility should be measured both in terms of human wellbeing and Gross Domestic Product (GDP) as it should be with any other technology. In this context, the knowledgeable part of the society, with sufficient expertise to design, manufacture, regulate and use the AI technology, will gain the most from it. Technologically less developed part of the society, with lack of digital expertise, will gain the least. If the GDP added value brought by the AI technology is fairly distributed for the wellbeing of the society the potential disparities may be mitigated.

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

Several organizations and communities are engaged to promote trust and understanding of AI. IEEE launched *The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems*⁶⁸⁴ (*The IEEE Global Initiative*) in April of 2017, and it is comprised of over two hundred and fifty global thought leaders and experts in AI, ethics, and related issues. The goal of The IEEE Global Initiative is to find broad consensus on how these intelligent and autonomous technologies can be aligned to moral values and ethical principles that prioritize human well-being. Public comments and input about the first version of their *Ethically Aligned Design* document received over one hundred and fifty pages of feedback from countries around the world, including China, Japan, India, Mexico, and Russia. The IEEE Global Initiative also identified many areas where standards are needed and, as a result, IEEE has initiated the IEEE P7000™ series of ethically oriented standards⁶⁸⁵.

Google, Facebook, Amazon, IBM and Microsoft have created the Partnership on Artificial Intelligence to Benefit People and Society⁶⁸⁶, dedicated to advancing public understanding of the sector of AI, as well as coming up with principles for future researchers to abide by. The Partnership aims to develop principles on ethics, fairness, privacy, trustworthiness, liability and safety. How ethics and values could be embedded into the AI algorithms; how to ensure transparency, fairness and accountability of the algorithms; whether there should be a general legal framework for algorithms or whether there should be sector specific regulations reason.

The public hearing on Artificial intelligence & Society⁶⁸⁷ (1 February 2017, Brussels) joined speakers from academic, corporate, and trade union backgrounds to discuss the broad impact of AI seen from all corners of society (labor, safety, privacy, ethics, skills, etc.). The input and information gathered at this hearing was intended to engage more stakeholders into defining a global policy on AI.

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

⁶⁸⁴ https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

⁶⁸⁵ <https://standards.ieee.org/develop/project/7000.html>

⁶⁸⁶ <https://www.partnershiponai.org/>

⁶⁸⁷ <http://www.eesc.europa.eu/?i=portal.en.events-and-activities-artificial-intelligence>

AI is both a standalone technology and an underlying component of many technologies, therefore many sectors stand to benefit of it. In particular, system industries (automotive, air and space, defense, energy industry, medical systems, manufacturing, transport) are going to be deeply changed by the surge of AI. Also, personalized medicine for early diagnosis, robotized surgery, AI-based healthcare, prediction and prevention of diseases, ecological and environmental disaster prediction, AI supported learning, smart cities, etc.

7. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Innovation-friendly regulation, based on standards, is required to cope with these problems. The regulation should address transparency and accountability of AI algorithms, risk management, data protection and safety. Certification of systems involving AI is a key technical, societal, and business issue. It should provide measures when someone (company, person) does something inappropriate and how to enforce the law. Good regulation should not stop innovation.

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

The 2016 IEEE AI & Ethics Summit⁶⁸⁸ (15 November 2016, Brussels) brought together technology leaders and policy makers to discuss their vision for what AI means to the future of humankind. The panels converged around the belief that prior to writing legal regulations, ethical issues need to be considered. How to program human ethics? Are the machines capable of making what humans consider as ethical or moral decisions? Should they make decisions or do humans need to be in the loop?

Civil Law rules on Robotics: Prioritizing Human Well-being in the Age of Artificial Intelligence, an event organized by Knowledge4Innovation and IEEE-SA and hosted by IEEE in the European Parliament in April of 2017, featured experts from The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, along with officials working on these topics, for a multifaceted dialogue. The event was hosted by MEP Mady Delvaux, who served as Rapporteur on the Civil Law Rules on Robotics Report. By addressing issues of autonomy and liability (including aspects of robotic personhood), the effects of

⁶⁸⁸ <http://ieee-summit.org/>

job transformations, and privacy and data protection, panelists explored how it is only by prioritizing human well-being when introducing AI into society that we will avoid unintended consequences and redefine progress in the age of AI.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

Lack of transparency should be acceptable when it addresses personal data protection. It should not be permissible regarding principles of human ethics and moral decision, safety and reliability of the AI system.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Any democratic government should take the leading role in setting a long-term AI strategy, instead of leaving it to industry and the research sector. This is particularly justified if the AI development is defined in terms of where we want to go, rather than how quickly we may get there⁶⁸⁹.

Major legal issues to be addressed by a new AI related legislation should include:

- Establish liability of industry for accidents involving autonomous machines (such as smart robots, driverless cars). This poses a challenge to existing liability rules where a legal entity (person or company) is ultimately responsible when something goes wrong
- New AI product safety regulations are needed
- Protection for citizens and businesses in case of malfunctioning software
- AI machines pose challenges in terms of data protection
- Mandatory insurance of AI products

Complementary to legislation, a guiding ethical framework for the design, production, and use of AI is required, based on the principles of human dignity and human rights, equality, justice, non-discrimination, and social responsibility. Ethical codes of conduct for AI researchers and designers, as well as licenses (rights and duties) for designers and users, need to be taken into account when proposing new legislation. Further to that, AI accountability and transparency have to be addressed as an explicit guiding ethical principle.

⁶⁸⁹ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2906249

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

The success of AI technology depends on the ease with which people use and adapt to AI applications. *Eurobarometer survey on autonomous systems*⁶⁹⁰ (June 2015, European Commission, Directorate-General for Networks, Content and Technology -DG CONNECT) looks at Europeans' attitudes to robots, driverless vehicles, and autonomous drones. The survey shows that those who have more experience with robots (at home, at work or elsewhere) are more positive towards their use. Moreover, the way AI systems interact with end users and help them to build cognitive models of their power and limits, are key technological objectives that help their adoption and sense of control.

Valuable recommendations to the EU policy makers regarding the consequences of Artificial Intelligence on the (digital) single market, production, consumption, employment and society are provided in the recently published report-opinion of the European Economic and Social Committee (rapporteur Catelijne Muller)⁶⁹¹.

6 September 2017

⁶⁹⁰ http://ec.europa.eu/public_opinion/archives/ebs/ebs_427_en.pdf

⁶⁹¹ <http://www.eesc.europa.eu/m?i=portal.en.int-opinions.40538>

IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems – Written evidence (AIC0100)

Introduction

It is with great pleasure that we submit information about our work as Officers of The IEEE Global Initiative in hopes it will be of immediate and pragmatic use to Lord Clement-Jones and The House of Lords Select Committee on Artificial Intelligence.

It is our distinct hope our work could be referenced, as multiple sections of *Ethically Aligned Design* and our IEEE P7000™ Standards Projects directly address the questions posted within your call for evidence. To that end, we have created the following document outlining the specifics of our work accessible via this link we'd like to submit in regards to your efforts:

[List of Accomplished and Ongoing Work by The IEEE Global Initiative](#)

Next Steps

We are delighted that Konstantinos Karachalios, Managing Director of The IEEE Standards Association, will be attending The House of Lords Select Committee on Artificial Intelligence's meeting on October 17th, 2017. The IEEE Global Initiative is technically a program initiated and supported by IEEE Standards Association, and Konstantinos is our greatest benefactor and supporter within IEEE as a whole.

As part of the panel on October 17, Konstantinos will be able to best position the work of IEEE and The IEEE Global Initiative within the context of the efforts of The House of Lords Committee. But beyond simply describing our work, it is the goal of IEEE and the Initiative to try and bridge the scientific and technical community developing these technologies with the political actors providing the legislation mirroring the issues of AI to society at large.

In this regard, IEEE is both a globally trusted, neutral player regarding the development and implementation of technology, and a convener and cross-pollinator who in this case can take the ideas stemming from this critical work from the House of Lords relevantly back to the vast, global engineering and academic community which IEEE represents.

Raja Chatila – Chair, The IEEE Global Initiative

Kay Firth-Butterfield – Vice-Chair, The IEEE Global Initiative

John C. Havens – Executive Director, The IEEE Global Initiative

Konstantinos Karachalios – Managing Director, IEEE Standards Association

5 September 2017

Imperial College London – Written evidence (AIC0214)

Submission to Select Committee on Artificial Intelligence

1. Imperial College London’s mission is to achieve enduring excellence in research and education in science, engineering, medicine and business for the benefit of society.
2. Artificial Intelligence (AI) research in the Department of Computing at Imperial College London is centered on the study and development of intelligent, autonomous systems.
3. Our research focuses on theoretical foundations as well as applications of AI. Our expertise ranges from machine learning to knowledge representation and reasoning, autonomous agents and multi-agent systems, human-machine interactions and collectives, cognition and human modelling, data science, robotics, augmented reality, graphics, computer vision and imaging, audio-visual signal processing, natural language processing and affective computing.
4. AI is currently experiencing a ‘spring’. This is mostly driven by recent successes of machine learning, a branch of AI, caused by the unprecedented availability of data and more powerful machines.
5. To maintain the momentum created by these successes they must be integrated with other forms of AI, and the resulting intelligent systems made to work in unison with humans to ensure this progress can deliver on its promise.

AI in the workplace: public perception, regulation and policy responses

6. The impact of machine intelligence on work is an issue that is already permeating the public consciousness, but eye-catching news story headlines on the issue may be inaccurately representing what the future of work looks like.
7. Replacement by machines emerged as a key concern from public dialogue exercises with Ipsos MORI⁶⁹².
8. Participants also questioned whether AI could potentially drive replacement of workers on a large scale – and across sectors – in a way that affected both skilled and manual workers. In contrast to technological changes in the past, which affected specific sectors, people see AI as driving more sweeping changes in how labour is organised. In tandem, participants were also concerned that increasing ‘intelligence’ could foster over-reliance on technology, with people de-skilling in certain areas –

⁶⁹² <https://royalsociety.org/~media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf>

for example, if medical professionals were relying on computers for diagnoses – as a result.

9. In fact, AI is more likely to mainly augment rather than replace jobs in the labour market. Individual tasks involving pattern recognition and repetitive actions may be automated, but the wholesale replacement of workers with machines in roles requiring thought, reason and relationship-building is improbable. The AI and computing science fields will need to more clearly explain the potential of AI to complement and aid existing roles, as well as creating new ones, to dispel commonly-held concerns about widespread job losses.
10. Regulation and informed certification of AI systems are important factors in public acceptance of AI. The whole field of formal modelling, verification measurement and performance evaluation of AI systems is still very much in its infancy: it is critical that one should be able to prove, test, measure and validate the reliability, performance, safety and ethical compliance – both logically and statistically/probabilistically – of such AI systems before they are deployed.
11. It should be noted that the verification of systems that adapt, plan and learn will involve the development of new modelling and verification approaches.
12. Insufficiently strong regulatory frameworks in emerging technologies could lead to societal backlash (not dissimilar to that seen with genetically modified food) should serious accidents occur or processes become out of control (such as algorithms used to determine a consumer's creditworthiness).
13. The potential development of increasingly autonomous artificial intelligence systems assessing job applications, controlling vehicles and weaponry poses serious ethical questions. One example that has seen some media coverage is how autonomous vehicles might decide to prioritise the lives of passengers over those of pedestrians⁶⁹³. A much greater emphasis needs to be placed on considering the ethical implications of automating some decision-making processes currently undertaken by a human.
14. Further ethical questions are posed by the ever-increasing use of human metadata that has driven the AI revolution. Preserving individuals' privacy whilst harnessing the potential of the ever-increasing data created by them living their technology-assisted daily lives – smartphone usage, medical records, satnav journeys – is a challenge that needs further exploration.
15. AI should enable developed economies such as the UK to become or stay competitive in a range of markets if it exploits the technology correctly. These

⁶⁹³ <https://www.technologyreview.com/s/542626/why-self-driving-cars-must-be-programmed-to-kill/>

technologies are becoming increasingly available to lower wage economies with whom the UK already competes as producers of knowledge based and other industrial products.

16. The UK suffers in labour-intensive industries because of wage costs, but automation can alleviate the gap with economies that have lower wages by empowering our skilled and semi-skilled workers to be more productive. This is also the case for “information-based” professions, such as legal and financial services, estate agents, financial analysts and traders who will increasingly need to rely on big data and machine intelligence.
17. However, the UK also needs to be aware that there is a risk that the cost gap between the UK and the lower cost economies may widen; this could happen if they are able to use smart automation and big data, including smart tutoring systems, to make up for their skill gaps and compete even more successfully with the UK.
18. Who benefits from AI-driven changes to the world of work will be influenced by the policies, structures, and institutions in place. Understanding who will be most affected, how the benefits are likely to be distributed, and where the opportunities for growth lie, will be key to designing the most effective public policy interventions to ensure that the benefits of this technology are broadly shared. To avoid creating a group of people who are left behind by the advance of this technology, action is needed to develop policy responses that will enable citizens to adapt to this new world of work.
19. At this stage, it will be important to allow policy responses to adapt as new implications emerge, and which offer benefits in a range of future scenarios. One example of such a measure would be in building a skills-base that is prepared to make use of new technologies, through increased data and statistical literacy.

Developing skills for AI

20. Education, at all levels (from primary school to university), needs to guarantee that every student leaves education data-literate. An ability to properly interrogate data and to understand bias in sample data will soon become an essential output of education, much as literacy and numeracy have been for decades.
21. Being able to programme a machine and understand algorithms is crucial in understanding what AI does and how to relate to and control it. Schools and (non-computer science) university courses are currently focusing on “computeracy”, namely IT (how to use machines). They instead need to focus on computer science and “programmacy” (how to program and control machines).

Human interaction with AI systems

22. Many AI systems are and will be deployed in situations where they interact with humans, or in settings where the data with which they interact is not static. This presents a number of technical challenges, opportunities and concerns, for example:
- How do we best combine human intelligence and artificial intelligence?
 - How do we ensure that AI systems will perform as expected with humans in the loop?
 - How do we design effective decision support tools based on AI?
23. Systems are already being developed to try to read the emotional state of a human, and respond accordingly in order to best address their needs. While this will provide many benefits, an important concern relates to how this may lead to undesirable influence of systems over humans. Even now, news sources and social media sites may tailor the stories and adverts that an individual is shown in order to maximise revenue. When these technologies are combined with perhaps more powerful, emotional channels, it is prudent to consider the effect on society. For example, one might imagine that data might show that angry people buy more of a certain type of product; this could lead a profit-maximising entity to promote anger-inducing stories. While such behaviour may already be part of our economic environment, AI-optimised channels of influence may raise the level of concern.
24. A related area of great interest is understanding how emotion helps humans to deal effectively with scenarios, and how these benefits might be incorporated into an artificial system. For example, through reinforcement learning, it is natural for an agent to weigh up the value of 'curiosity' when considering whether to invest additional resources into exploring potential actions and their consequences.
25. In addition to these areas of human-computer interaction, there are areas of research to explore how humans and AI systems can work together in partnership. To create effective partnerships, it is necessary to understand the strengths of each partner, and design systems with these in mind.
26. Can we create AI systems whose workings, or outputs, can be understood or interrogated by human users, so that a human-friendly explanation of a result can be produced? Increasing the interpretability of machine learning methods is desirable for a number of reasons, as noted earlier. These include the need to understand the processes used in safety critical systems or the ways in which decisions about individuals have been reached. There are different possible approaches to achieving interpretability.
27. AI methods could be restricted to those that directly yield an interpretation which is easy for humans to understand. One example of such an approach is a decision tree, which repeatedly makes sequential decisions according to simple rules. However, a significant drawback to this approach is that there may be important trade-offs between interpretability and accuracy. Further, if only repeated simple decision rules

are allowed, then in order to make accurate predictions, it may be necessary to apply many thousands of rules, thereby losing the desired feature of interpretability. A more nuanced approach involves tackling a classification or prediction task as a pipeline of machine learning models. The output from each model is fed as input into the subsequent model, such that the detection of generic features in the original input and formulation of a classification or prediction based on these features is split into two or more stages. The benefit of this approach is that the intermediate outputs of models in the pipeline can be designed so as to be interpretable by humans.

28. A more elaborate approach would be to create an interface between AI systems and human-machine dialogue systems so that, in the future, humans could talk to the machine and interrogate its reasoning. This may seem very appealing, but it will rely on underlying explanatory ability combined with confidence in the speech interface, where ambiguities might creep in and lead to potential misunderstanding.

Supporting AI research and innovation

29. There is real need for fundamental research, particularly in various aspects of AI such as planning, perception, language understanding/generation, reasoning, multi-modal information fusion, modelling the human, and dealing with the uncertainty present in the real world in which AI systems must operate in general.

30. Furthermore, very few current AI systems, even those which display impressive performance such as AlphaGo, or the growing generation of autonomous vehicles which do operate in the real world, really understand their own reasoning processes – let alone have the ability to explain their reasoning and be aware of the limitations.

31. It is critical both for understanding the reproducibility of such systems and for their use in important applications such as Decision Support Systems for critical applications, including human health and safety, that research in this area is promoted. Support for this research, such as the recently announced EPSRC priority area of Human-Like Computing (2016-2020) is welcome.

32. Human-Like Computing (HLC) research aims to endow machines with human-like perceptual, reasoning and learning abilities which support collaboration and communication with human beings. Such abilities should support computers in interpreting the aims and intentions of humans based on learning and accumulated background knowledge to help identify contexts and cues from human behaviour. The development of computer systems which exhibit truly human-like learning and cooperative properties will require sustained interdisciplinary collaboration between disparate and largely disconnected research communities within Psychology and Artificial Intelligence.

11 September 2017

Information Commissioner’s Office – Written evidence (AIC0132)

Contents

Executive Summary

Introduction.....

Historical Context

The Pace of Technological Change

Impact on society

Public Perception

Industry

Ethics.....

The role of the Government

Learning from others

Annex- Basic Compliance Steps for the responsible use of AI

Executive Summary

1. The Information Commissioner's Office (ICO) is responsible for promoting and enforcing the Data Protection Act 1998 (DPA) and welcomes the opportunity to respond to the Committee's Call for Evidence.
2. Our interest in Artificial Intelligence (AI) technology lies in the processing of personal data. The automated processing of personal data without appropriate checks has been a privacy concern for many years. Successive data protection laws including the current DPA, and the EU General Data Protection Regulation coming into effect in May 2018 require organisations who process personal data to comply with a number of important principles. These principles help to mitigate privacy risks and provide certain rights to individuals.
3. The rapidly increasing use of AI, although a type of automated processing, presents its own unique risks. Whereas 'traditional' processing involves a human-being making decisions as to how and for what purpose data is processed, AI enabled processing involves a computer making these decisions with little or no human oversight. There are questions to be answered as to whether a computer can show the same level of empathy and reasonableness in making often significant decisions about individuals.
4. People have mistrust in the use of AI technology. We believe the key elements for preparing the public are: transparency- providing individuals with information about the implications and likely outcomes from the use of AI; control – ensuring a significant element of human oversight and intervention through knowledgeable, appropriately senior, dedicated staff; and effective regulatory oversight – organisations taking a number of compliance steps including regular reviews and privacy impact assessments.
5. The use of AI raises ethics as well as privacy concerns. Data protection law, especially new requirements contained in the soon to be implemented General Data Protection Regulation, go a long way in tackling these concerns. Ultimately, data protection is about the relationship between those that process personal data and the people whose data is being processed. If those who use AI do so fairly then many of these concerns about its use will be addressed. This will be to the benefit of impacted individuals and society as a whole.

Introduction

6. The Information Commissioner's Office (ICO) has responsibility in the UK for promoting and enforcing the Data Protection Act 1998 (DPA), the Freedom of Information Act 2000 (FOIA), the Environmental Information Regulations 2004 (EIR), the Privacy and Electronic Communications Regulations 2003, as amended (PECR), and the eIDAS Regulations (2016). We also deal with complaints under the Re-use of Public Sector Information Regulations 2015 (RPSI) and the INSPIRE Regulations 2009. We are independent of Government and uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals. We do this by providing guidance to individuals and organisations, solving problems where we can, and taking appropriate action where the law is broken. We welcome the opportunity to respond to your call for evidence and are grateful for your consideration of this submission.
7. The Information Commissioner's interest in artificial intelligence (AI) lies primarily where its use involves the processing of personal data. Personal data is defined in the DPA as "data which relate to a living individual who can be identified from those data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller."⁶⁹⁴
8. The processing of personal data by automated means has always posed a privacy risk for individuals. Since the UK's first Data Protection Act in 1984, successive legislation has required organisations to take certain steps to mitigate these risks whilst giving individuals' specific rights over their own personal data.
9. AI, although a type of automated processing, creates its own unique risks that are potentially even more intrusive to individuals' privacy. Unlike other forms of automated processing, AI programs don't linearly analyse data in the way they were originally programmed. Instead they learn from the data they have already analysed in order to respond intelligently to new data and adapt their outputs accordingly. This brings the possibility of AI-enabled technologies making significant decisions about people, with little or no human oversight. This evidence makes clear that data protection rules have become more relevant than ever, and if applied effectively can help to protect individuals, mitigate risk and to allow society to reap the benefits of AI technology.

⁶⁹⁴ Data Protection Act 1998, S1 (1)

Historical Context

10. The use of automated data processing without appropriate checks and balances has been a privacy concern for many years. The OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980) represent one of the earliest international data protection instruments. The Guidelines' Explanatory Memorandum states:

"As far as the legal problems of automatic data processing (ADP) are concerned, the protection of privacy and individual liberties constitutes perhaps the most widely debated aspect. Among the reasons for such widespread concern are the ubiquitous use of computers for the processing of personal data, vastly expanded possibilities of storing, comparing, linking, selecting and accessing personal data, and the combination of computers and telecommunications technology which may place personal data simultaneously at the disposal of thousands of users at geographically dispersed locations and enables the pooling of data and the creation of complex national and international data networks."

11. Subsequently, the current European Data Protection Directive (95/46/EC), adopted in October 1995 - which still forms the basis of the UK's current data protection law - states in its second Recital:

"Whereas data-processing systems are designed to serve man; whereas they must, whatever the nationality or residence of natural persons, respect their fundamental rights and freedoms, notably the right to privacy, and contribute to economic and social progress, trade expansion and the well-being of individuals."

It's worth pointing out that the OECD Guidelines and the Directive 95/46/EC were drafted during an era of stand-alone computers and basic telephony systems with very limited functionality. The internet was not widely used in business or for personal use. Most of the technology companies whose services we are so dependent on today had not yet been founded. The roots of social media, wide-spread data-sharing, Big Data and artificial intelligence (AI) were just forming. Therefore we believe that both the OECD Guidelines and the Directive were highly prescient, and were right to acknowledge the threats as well as the opportunities of information technology. At the time of drafting, to many, the risk of mankind serving technology, and not vice versa, must have seemed the stuff of dystopian science fiction. However, the use of AI has the potential to bring this risk closer to home.

12. DP laws have recently been modernised to tackle the challenges of technology in the twenty-first century. The EU General Data Protection Regulation was passed in 2016 and will come into effect in May 2018. A UK Data Protection Bill will also be introduced to cover national implementing measures and areas where member states are allowed to derogate. The legislation increases individuals' rights – for example rights in relation to profiling and introduces new concepts such as data protection by design and data protection impact assessments.

The Pace of Technological Change

13. Question one. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

14. There will be others who are directly involved in technical development who may be better placed to comment on the technical factors affecting development in this area or how it is likely to develop over the next few years. However, we are aware of a general increase in the adoption of AI technology and how swiftly this is becoming a mainstream technology, with a wide range of potential uses in both the public and private sectors. The volume and range of datasets available, increases in computing power and online storage are rapidly driving forward these advances. The Information Commissioner published a report on the implications of AI for data protection earlier this year⁶⁹⁵. AI will also feature as a priority area in the Commissioner's new Technology Strategy, which will be published later in 2017.

15. A lack of public trust could be a factor that hinders the take-up of AI, particularly in personal data processing contexts. ICO research conducted in 2016 found that only one in four UK adults trust businesses with their personal information.⁶⁹⁶ Trust may be even more lacking when it comes to the use of AI and automated processing more generally. There could be a point at which public suspicion – arising from a lack of control and understanding –

⁶⁹⁵ <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

⁶⁹⁶ <https://ico.org.uk/media/about-the-ico/documents/1624382/ico-annual-track-2016.pptx> p.10

undermines trust and inhibits the take-up and development of new services – particularly digital ones.

16. The lack of public trust could be compounded by the perception that decisions based on AI are opaque at best and have unfair or otherwise undesirable consequences for people. Questions arise such as what criteria is the computer using to carry out certain actions? How will I know if the 'computer is wrong'? And what can I do about it? A useful automated processing albeit not AI example is that of the Border Systems Programme use for security purposes. There are public perceptions that the system is setup to target individuals on the grounds of race or religion, where in fact this is a misunderstanding. A lack of information as to how the system works likely contributes to this mistrust.

17. The responsible use of AI for the electronic delivery of government services is very important – of course we want the public to provide accurate data and to take up willingly the new services that technology facilitates. This depends on the transparency, control and oversight that we will elaborate on later in our evidence.

18. Recent UK research⁶⁹⁷ found that 55% of UK consumers find AI 'creepy'. However, as we explain later in our evidence, there are generally ways of mitigating the risks and of keeping this mistrust at bay. However, it is possible that some uses of AI will always be unacceptable. Should an individual's innocence or criminality ever be automatically inferred using solely automated / AI means?⁶⁹⁸ If so, where should the legal and ethical limits to the deployment of such technology lie?

19. Question two. Is the current level of excitement which surrounds artificial intelligence warranted?

⁶⁹⁷ <https://www.research-live.com/article/news/half-of-uk-consumers-find-artificial-intelligence-creepy/id/5024372>

⁶⁹⁸ <https://arxiv.org/pdf/1611.04135v1.pdf>

20. The Information Commissioner believes that the current level of excitement is warranted but needs to be tempered by caution and a comprehensive assessment of the risks and benefits. There are certainly significant benefits to the use of AI but there are also data protection implications. We discuss this in detail in our research paper.⁶⁹⁹
21. The use of AI presents some novel challenges for data protection safeguards. A classical paradigm in data protection is where a data controller (the person, usually an organisation, who decides the purpose for and manner in which the data is processed) processes information about an individual for a particular purpose – for example to work out a housing benefit claim. A human being will work out what information is necessary to process the claim, where it should come from and how it should be analysed to produce the right result. The technology used will be essentially inert and will only process the information in the way it is programmed to. Once the claim is processed, a human being will deal with any ensuing disputes or queries, hopefully using the very human principles of reasonableness, fairness and with perhaps an element of empathy.
22. An AI-enabled scenario can be different to the above scenario in several key respects. Whilst issues of data controller responsibility and purpose might be the same, decisions over the sources of the data and the methods used to analyse it could be taken by the AI-enabled devices themselves. This in turn has implications in terms of compliance with other data protection rules, such as transparency, fairness, necessity, relevance and adequacy. Although AI is reportedly becoming more intelligent there are also issues over whether a machine could really display the reasonableness and empathy that can be needed to deal with individuals. This illustrates the importance of human supervision and intervention when AI is in use – we discuss this in greater detail below.

Impact on Society

23. Question three. How can the general public best be prepared for more widespread use of artificial intelligence?

⁶⁹⁹ <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

24. It is clear that the use of AI is increasing and is being used to make decisions that can have significant impact on people. Examples of this include the use of AI in internet counter-terrorism surveillance, offender management, credit referencing and on-line dispute resolution. It seems likely that the use of AI to analyse personal and non-personal information will continue to expand into more areas and to have a more significant impact on peoples' lives.
25. In our view the key elements for preparing the public for the more widespread use of AI are transparency, control and effective regulatory oversight. We consider each of these elements in turn.

Transparency

26. It is a basic and crucial requirement of data protection law that – in normal circumstances – people should be aware of such matters as who is collecting their information, how it will be used and whether it will be disclosed to a third party. This information is usually communicated to the public through an organisation's privacy notice. Even where a data processing operation involves the use of AI it should still be possible to provide this basic privacy information. However, current data protection law also contains provisions intended to protect individuals against the potentially negative impact of automated decision-making, including the use of AI.
27. The General Data Protection Regulation (GDPR), which will be implemented in the UK in May 2018, places more emphasis than the current law on automated decision making, when used for purposes such as profiling an individual – for example to target behavioural advertising.
28. In terms of transparency, in certain circumstances, the legal requirement under the GDPR will be for individuals to be made aware that automated decision making is taking place, to be provided with *meaningful information*

about the logic involved, as well as the significance and the envisaged consequences of the data processing. This is where the use of AI as part of a personal data processing activity poses real challenges in terms of transparency and intelligibility to the public; what is meaningful information about AI? The problem is that the 'math' behind the algorithms used in AI would only be understandable to a limited number of experts and it would be very difficult for the vast majority of members of the public to challenge an AI-supported automatic decision on the grounds that its outcome is unwarranted, unfair or otherwise detrimental.

29. Individuals not being able to challenge such decisions would mean that – as the use of AI develops – there is the possibility of a widening rift of understanding between the public and the organisations that are using AI to make decisions about them. It may be more realistic for individuals to be provided with information about the implications and possible outcomes of the AI, rather than detail of the algorithm itself. In reality, it could be very difficult for members of the public to exercise their legal rights – and to be protected from the possible excesses of AI – without some form of expert mediation – we discuss the form that this might take in our comments on regulation below. Transparency will remain important but must be complemented by other effective safeguards.

Control

30. The second main right that individuals enjoy in respect of automated decision making, including the use of AI, is the right not to be subjected to a solely automated decision making process if the decision has 'legal' or a 'similarly significant' effect on an individual. The relevant provisions in the GDPR are complex, but in certain circumstances the individual also has a right to have an automated decision subjected to human scrutiny, to express his or her point of view and to contest the decision.

31. It is important to be aware, however, that it seems likely that organisations such as large e-commerce sites that use AI for purchaser – vendor dispute resolution will deal with a very large number of cases. It could be a challenge therefore for companies like this to offer complainants a second decision, taken using human intervention. There are also issues around how these – and other companies make individuals aware that AI is being used. Clearly

individuals cannot use their 'automated decision making' rights unless they know automated decision making – possibly involving AI – is in use.

32. The rights in data protection law are potentially very powerful in respect of AI-based decision making. They would mean, for example, that if a Credit Reference Agency (CRA) recommends that a credit grantor turn-down an application for a loan – based on an automated decision – then the person applying for credit would be able to contact the CRA if he or she considers the decision to be unfair, and the CRA would have to ask a 'real person' to re-assess the factors that were used to make the original decision; of course the outcome might be the same. However, the point is that the law recognises the risks that AI and automated decision-making can pose and gives people a 'human defence' against this.
33. It is worth noting that the ICO currently receives very few complaints about AI or automated decision-making – this suggests that on the whole these technologies are being used responsibly and with reasonable outcomes for individuals. (Or, on the other hand this could be the result of a lack of public awareness.) However, we expect complaint numbers and volumes of queries to rise as the use of AI becomes more prevalent and moves into potentially more controversial uses of data – for example using social media data to predict an individual's credit score.
34. We believe that the element of human intervention addressed above is particularly important in the specific context of AI. A unique aspect of AI is that algorithms can 'teach themselves' and develop, based on their 'experience' performing a particular task. This can of course have positive social consequences – for example an algorithm used to select particular travellers for counter-terrorism checks at airports – could become more accurate in the light of experience, leading to fewer false-positives and minimising collateral privacy damage. However, there is a danger that as technology 'makes its own rules' the results of its use could deviate from intended outcomes.
35. It is very important that organisations using AI applications review periodically the consequences of their use on the individuals whose data they are analysing and ensure the processing activity has not deviated from its intended purpose and is not having unintended consequences. Ensuring that

organisations have rigorous processes underpinned by knowledgeable dedicated staff, including data protection officers with the correct level resources and organisational influence, will also be important.

Regulation and organisational accountability

36. As we have explained above, current data protection law and the GDPR both contain features that are highly relevant to the use of AI in personal data processing contexts. Appropriate use of AI for the processing of personal data depends on going through a series of compliance steps. We have included a list of such steps as an annex to this submission. Such checks, however, should not be viewed as a 'tick box' process and should be considered as comprehensively as possible likely through the use of ongoing privacy impact assessments.

37. As noted above, the requirement for organisations to be transparent about the uses of AI will have limitations for individuals. This therefore highlights the importance of ensuring organisations are accountable for their use of AI to data protection authorities. The concept of algorithmic accountability is important - organisations will need to provide evidence of how they have assessed and audited the impact and effects of the AI they have deployed. This may require the development of automated tools and new audit methodologies.

38. The Commissioner recognises that she will need to recruit more technical experts to audit and investigate issues related to AI. It will also be important that the market provides more services that audit AI - this also fits with the concept of 'certification' in GDPR - where the Commissioner will be able to accredit expert third parties to provide data protection certification that demonstrates compliance with the law.

39. The Information Commissioner recently completed an investigation into the trial of a service provided by the AI Company, Google Deepmind, to the Royal Free Hospital⁷⁰⁰. She concluded that the Hospital breached the Data Protection Act and required an undertaking to be signed to address non-

⁷⁰⁰ <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/>

compliance. The findings highlighted the importance of transparency, rigorous privacy impact assessment, robust contractual arrangements to prevent the re-use of patient data and verifying processes in practice using audits.

40. Despite robust data protection compliance, the law only takes us so far. We believe that it can be highly challenging to apply certain data protection concepts such as fairness and relevance to advanced AI applications. For example, empathic computing involves the use of AI to examine an individual's on-line behaviour. It considers the vocabulary individuals use, the way they input type and the pictures they look at longest in order to assess that individual's mood and deliver content accordingly. This certainly involves the processing of personal data and therefore engages data protection law. However, whilst the pure data protection compliance aspects of using AI in empathic computing and other contexts can be addressed using the compliance steps outlined in the annex, the use of AI raises wider ethical issues of significant public interest.

1.

41. Data protection law deals well with data processing activities – including those using AI – when the information being processed is about individuals and has an effect on those individuals. However, the broader social effects of technology, including AI, go beyond this. The creation of a data ethics advisory body may be a means to help ensure the public is engaged in these ethical issues. It would act to monitor the effects of technology on society, engaging with the public and providing advice to existing regulators to help ensure that the balance between the power of technology – and those controlling it – and wider societal concerns including the rights of individuals is struck correctly. The Information Commissioner is keen to ensure the right solutions are in place and is working with government to help with its consideration of the issue. It is important to ensure that any new advisory body would complement the existing work of the Information Commissioner and other regulators rather than seek to replace existing functions.

42. A data ethics advisory body's role should involve identifying data-related problems that existing regulators may not be able to counter, because they are unaware of them or because the problem falls outside their area of statutory competence. It could detect areas where the societal advantage of data use (personal or non-personal) is not being gained because, perhaps, of a misunderstanding or lack of relevant law. In such cases, a data ethics

advisory body could invite the appropriate regulator – or regulators - to provide clarification or make recommendations for law reform.

43. Question four. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

44.This question is not applicable to the Information Commissioner.

Public Perception

45. Question five. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

46.The situation with AI is much the same as with other technologies. Most people would probably be unable to explain what the main components of a computer do or how coding works. Nonetheless, they may be able to use a computer and understand the consequences of their digital activity.

47.Ideally members of the public would understand what artificial intelligence is and how it affects them. However, we need to be realistic about the public's ability to understand in detail how the technology works. Perhaps it would be better to focus on the effect of the technology – in terms of benefits and detriments – and to ensure that there is an effective regulatory system which does have the necessary technical understanding in place.

48.As we have explained elsewhere in our evidence, even though the 'math' may be difficult for non-experts to understand, it ought to still be possible to explain the purpose(s) for which peoples' data is being processed, who is doing the processing and the consequences of this. If we focus on the consequences of AI, rather than on the way it works, then it is possible to bring about public understanding and to allow individuals to exercise their rights.

49. The ICO has produced a new code of practice on privacy notices⁷⁰¹. This code stresses the need to communicate with the public in clear, accessible ways. The guidance in the code of practice is as applicable to data processing carried out using AI as it is to more conventional forms of data processing.

Industry

50. Question six. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

51. This question is not applicable to the Information Commissioner.

52. Question seven. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

53. Please refer to our evidence above.

Ethics

54. Question eight. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

⁷⁰¹ <https://ico.org.uk/for-organisations/guide-to-data-protection/privacy-notice-transparency-and-control/>

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

55. We have already addressed most of these issues earlier in our evidence. However, we would like to clarify that data protection law is framed in terms of the relationship between data controllers (organisations) and data subjects (individuals). However, it could also be seen as being about the relationship between individuals (acting on behalf of organisations) who make decisions about information use and individuals who are affected by those decisions. In that sense, data protection can be seen as a branch of ethics. We believe that many ethical issues and questions relating to societal norms will be addressed provided that the relationship between organisations and the people whose data they analyse, whether or not using AI, is a fair one.
56. Utilising modern data protection regulatory concepts will also be important. These include: ensuring technological capabilities are used in a proactive way to safeguard privacy- privacy by design; the impacts are understood and addressed at the outset- privacy impact assessments; and that organisations take proactive responsibility once processing is underway- accountability.
57. Other risks, such as diversity, also highlight the importance of organisations undertaking privacy impact assessments and broader ethical impact assessments before commencing the implementation of AI. Recent research highlights the risks that AI can pose for gender and ethnicity issues⁷⁰².
58. The Information Commissioner recognises the importance of applied research that considers the risks of AI but also looks for innovative privacy enhancing solutions that can make a real difference to the public. Her recent Grants Programme encouraged applications in relation to AI. 119 applications have been received for the programme and the grants awarded will be announced before the end of the year⁷⁰³.

Consent

⁷⁰² <http://www.sciencemag.org/news/2017/04/even-artificial-intelligence-can-acquire-biases-against-race-and-gender>

⁷⁰³ <https://ico.org.uk/about-the-ico/what-we-do/grants-programme/>

59. We would like to add a comment about consent. The role of consent is often misunderstood – it can be seen both as a cure-all and as a legal requirement of data protection law. For the reasons we have already discussed, there can be real problems in expecting people to consent to their data being processed by AI systems. Many people will not know what AI is or the implications of its usage, and in data protection law consent has to be fully informed to be valid. This means that there may be significant problems in legitimising the use of AI on the basis of individuals' consent.
60. Individuals can suffer from 'consent fatigue' – as may be the case with repeated 'cookie consents'. Many individuals might prefer the services and systems they use to do what they expect them to and to use their personal data fairly and responsibly, with benign and predictable outcomes, but not to be repeatedly asked for their consent. Another problem is that if the use of AI is occurring on the basis of consent, then it would likely have to cease if consent is withdrawn or found to have not been given in the first place. This could lead to the scenario of organisations offering AI and non-AI enabled services, something that it could be infeasible to deliver in practice.
61. On a legal point, data protection law is sometimes portrayed as requiring individuals' consent in order to process their personal data. This is not the case. The law usually provides a number of bases for processing personal data, consent is just one. Organisations can process personal data, including the use of AI provided the activity is legitimate and does not have a detrimental effect on people. If this is the case and the compliance issues we have discussed earlier in our evidence are addressed properly, then organisations should be able to go ahead with the processing without the individual's consent. It is important to be clear, however, that this is not the case with regards to the processing of 'sensitive' personal data – for example, data relating to the health, racial or ethnic origin, political opinions or sexual orientation of individuals. Here, consent or another appropriate basis will need to be used.
- 62. Question nine. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?**

63. As we have already explained, transparency is one of the basic requirements of data protection law. However, there are exemptions from this, for example where providing too much information about how a system operates would prejudice the purposes of law enforcement, or where providing information about the logic involved in decision-taking constitutes a trade secret. The rules and norms here are well-established and apply equally to AI and non-AI processing of personal data.

64. Regardless as to whether an exemption does or does not apply, there is still a concern with 'blackboxing' in terms of accountability. An issue with the 'black-box' is that no-one understands how an AI system got from input to output. Where there is zero transparency how can the processing be demonstrably compliant with data protection laws? We discuss potential methods to approach algorithmic transparency in our research paper.⁷⁰⁴

The Role of the Government

65. Question 10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

66. Government must recognise that there are unique features of AI that mean that it presents risks as well as opportunities. We do not think that AI should be regulated as a discrete topic. We should look more at the purpose and effects of its use, rather than the technology itself. As we have already explained many uses of AI are already subject to regulation through data protection and other laws. However, as discussed above, given the complexity of the regulatory landscape and the fact that AI straddles several areas of regulatory responsibility, we do think there is a case for some form of ethics advisory body to take a holistic view, providing advice to existing regulators so that the best protection is offered to individuals and to society as a whole.

67. We should not underestimate the potential consequences of AI for individuals, ones that can be irreversible. This is why it is so important that organisations deploying AI-enabled systems have a clear set of compliance rules so they

⁷⁰⁴ <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

can design and deploy AI systems properly, with proper respect for the individuals whose data they may be processing. We have explained earlier on in our evidence how data protection legislation provides appropriate safeguards where personal data is involved.

Learning from Others

68. Question 11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

69. The International Conference of Data Protection and Privacy Commissioners, the forum for the world's data protection and privacy authorities of which the ICO is member, focussed specifically on the topic of AI as part of its 38th gathering in 2016. The fact that this theme was chosen for the conference demonstrates the significantly increased level of global attention that AI devices have attracted in the last two to three years. There is consensus across data protection and privacy commissioners that we are only just beginning to understand the challenges that AI brings to data protection. The Information Commissioner will continue to work with her international counterparts in furthering the understanding of these challenges and proposing potential solutions.

Annex- Basic Compliance Steps for the Responsible use of AI

70. These are the basic steps that should be taken when implementing an AI-enabled data processing system. They are based on the premise that the use of AI is going to become more prevalent and that organisations need to understand the rules needed to deploy it responsibly.

- 1) Initial assessment of the need for the data processing; what are you trying to achieve – e.g. detect fraudulent benefit claims and why is AI based processing a necessary and proportionate response to achieving this? (Commissioning a form of privacy impact assessment may help with this and to identify necessary conventional data protection safeguards if personal data is involved).
- 2) If the decision is taken to use AI, specify a range of data inputs (i.e. data sources and data items) as well as limits on algorithmic self-improvement.
- 3) Test the system, ideally using synthetic data or, if this is not possible, a small sample of live data (in accordance with appropriate safeguards) and assess the results – e.g. is benefit fraud being detected accurately?
- 4) If the system is intended to go live, ensure that during the design and testing phases, transparency procedures for informing the public of general privacy information but also of the use of automated decision making / artificial intelligence are developed.
- 5) Carry out regular audits to ensure that the system is working in the expected manner - i.e. that the correct data items are being utilised and that they are being analysed in accordance with design parameters.
- 6) Put systems in place for the periodic review of outcomes – is the system continuing to achieve its intended objectives? If not, modify the system or deploy a better one.
- 7) Ensure there are procedures in place for dealing with queries and complaints from the public, including means of re-taking a decision with an element of human intervention, and for delivering all relevant individuals' rights.

6 September 2017

Information Systems Audit and Control Association (ISACA) London Chapter – Written evidence (AIC0193)

The Select Committee on Artificial Intelligence appointed by the House of Lords
*Call for Comments on the economic, ethical and social implications of advances
in artificial intelligence*

Background and Summary

1. This submission is made by the ISACA London Chapter (ILC) in alignment with its parent organisation headquarters in the U.S., in response to the consultation call by the Artificial Intelligence (AI) Committee of the UK House of Lords, 29 July 2017.⁷⁰⁵
2. It covers key assumptions on AI, inclusive of the challenges said assumptions bring up, and practical solutions that we envisage can support the successful adoption of AI initiatives.
3. As an independent, non-profit, global association, ISACA engages in the development, adoption and use of globally accepted, industry-leading knowledge and practices for information systems.
4. Previously known as the Information Systems Audit and Control Association, ISACA now goes by its acronym only, to reflect the broad range of IT governance professionals it serves.
5. ISACA provides practical guidance, benchmarks and other effective tools for all enterprises that use information systems. Through its comprehensive guidance and services, ISACA defines the roles of information systems governance, security, audit and assurance professionals worldwide. The COBIT framework and the CISA, CISM, CGEIT, CRISC and CSX certifications are ISACA brands respected and used by these professionals for the benefit of their enterprises.

Methodology and Definition

6. The following section of the submission covers our approach and definition of AI:

⁷⁰⁵ <http://www.parliament.uk/documents/lords-committees/Artificial-Intelligence/Artificial-Intelligence-call-for-evidence.pdf>

7. In preparing this submission, the Government and Regulatory Advocacy Committee of the ILC conducted a preparatory phase by requesting members via the ILC newsletter to indicate interest to contribute their expertise to this submission, liaising with ISACA HQ in the U.S. for chapter and headquarter alignment of objectives; and using material on AI recently published in the ISACA Journal and Tech Briefs which draws on a pool of expertise worldwide. This submission therefore benefits from ISACA efforts and expertise among its membership which globally comprises of the following sectors in terms of employment, 66% of which not only influences large swathes of the digital economy, but also could be expected to be influenced by developments in Artificial Intelligence:

Technology Services/Consulting	31.4%
Financial/Banking	21.3%
Government/Military	7.7%
Public Accounting	6.5%
Healthcare/Medical	2.8%
Pharmaceutical	0.8%

8. **Definition.** It is commonly reported that definitions about AI vary and use of the term is often 'confused' with other related terminology such as automation and machine learning. At the heart of AI is its resemblance to human behaviour with some experts indicating that 'true' AI will not ever be possible if such technologies are to replicate all of human behaviour, including emotions such as empathy and trust. However, use of technologies that incorporate many human-like processes are deemed to provide a transformation in the business, government and consumer world.

9. AI defined in this submission therefore, reflects a two-step process to facilitate a distinction to illustrate what AI in its partial forms can achieve and how it differs from other non AI processes such as automation and forms of machine learning. In 2015, ISACA defined machine learning as:

*"The use of computing resources that have the ability to learn (acquire and apply knowledge and skills that maximize the chance of success). These cognitive systems have the potential to learn from business related interactions and deliver evidence-based responses to transform how organizations think, act and operate."*⁷⁰⁶

⁷⁰⁶ www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/innovation-insights.aspx

In 2017, ISACA took this definition of machine learning a step further to define AI:

"AI includes not only algorithms and technologies that learn and adapt in response to success or failure (or react to context), but also technologies that resemble human behaviour in the ways that process input and generate output."

Questions

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this?

We have only just started to see the possibilities of what artificial intelligence can do for society. The potential for good and bad is in our hands. Artificial intelligence (AI) enables organizations to attribute meaning to, understand the nuances of, and derive insights from data they may already collect via standard business processes. AI may help automate tasks that historically were completed by humans with specialized knowledge and skills. "We see this in the field of medicine, for example, where AI is helping to diagnose and treat patients with cancer and heart disease at record rates. We see AI working within the financial world as well, where the traditional Wall Street trading model is changing from traders on the floor to enterprises like Sentient, Wealthfront, Two Sigma and many more leveraging engineers with graphics cards and server racks to complete a deal."⁷⁰⁷

AI in practical terms can be viewed as *"the convergence of machine processing, learning and control"*⁷⁰⁸

In research consultancy Gartner's 'Top 10 Strategic Technology Trends for 2007' survey, Gartner Vice-President and Fellow David Cearley said

2. *"over the next 10 years, virtually every app, application and service will incorporate some level of AI. This will form a long-term trend that will continually evolve and expand the application of AI and machine learning for apps and services."*

How is it likely to develop over the next 5, 10 and 20 years?

We see it in the news often, AI is threatening jobs. "A report by PwC says that more than 10 million UK workers are at high risk of being replaced by robots

⁷⁰⁷ http://www.huffingtonpost.com/entry/artificial-intelligence-might-overtake-medical-and-finance-industries_us_599b201de4b0771ecb064fb6

⁷⁰⁸ <https://www.icaew.com/en/technical/information-technology/it-faculty/chartech-magazine>

within 15 years as the automation of routine tasks gathers pace in a new machine age.”⁷⁰⁹

In reality, on certain tasks AI may eliminate the need for human labor and require us as a society to think differently about how we work and workforce needs. Planning for the next generation workforce will indeed be critical and flexibility among those in the workforce important. However, AI will also bring efficiency and in many cases, remove human error from the equation as well as allow professionals to focus on other areas of the business that provide additional value.

The key will be thinking differently about how we work. AI represents another shift in the job market—like we saw with the manufacturing boom, the computer generation and now the advent of AI. AI solutions, services and systems will impact industries across the globe and across the spectrum, from transportation to manufacturing to healthcare and finance and yes, even government and military, with the potential to do much good.

What factors, technical or societal, will accelerate or hinder this development?

Fear of change will hinder development. There must also be the level of expertise available to design, implement, support and maintain the technology.

The biggest challenges, however, may be more societal than technical. We have already seen the disruption in labour markets that increased automation has brought about. The automation evolution began, and public policy did not keep pace; there were some efforts to reconfigure things like education, social programs, and the like, to better assist those displaced by automation in finding a new role in a changing workforce. The changes AI will bring, however, will make the automation evolution pale by comparison. It is therefore even more critically important that social considerations of the impact of AI begin to take center stage, and that public policy, regulatory measures, and international norms begin to evolve as well. A professional workforce unable or unwilling to take part in an AI-augmented workforce of the future will hinder AI development dramatically.

To accelerate AI’s development, these societal concerns must be addressed while simultaneously creating an environment in which AI can grow. This means stable and adequate funding for R&D, an increased use of AI within areas such as personalized learning and predictive analytics, and similar areas. Providing incentives to the business and academic realms to support the growth of AI will not only accelerate the development of AI products, services and solutions, but will likely generate the by-product of solutions to some of the societal concerns that the increased presence of AI in the workforce will bring with it.

⁷⁰⁹ (<https://www.theguardian.com/technology/2017/mar/24/millions-uk-workers-risk-replaced-robots-study-warns>).

3. Is the current level of excitement which surrounds artificial intelligence warranted?

Yes. The potential for good is enormous with artificial intelligence. This is life-saving technology, as medical science has shown us, creating machines helping to diagnose disease. As people become more used to AI technology, they will come to expect the real-time personalized services the technology provides. In terms of perceived impact and influence, Klaus Schwab, Founder and Executive Chairman of the World Economic Forum, states:

*"We're in the midst of a "Fourth Industrial Revolution", after steam power (the first), electric power (the second) and digitization (the third). The fourth, which incorporates AI and robotics as well as other technologies, will have an even greater impact"*⁷¹⁰

More recently, as the development of AI becomes an increasing national security concern, world leaders have expressed a growing importance in the sector. In an address to students on 1 September 2017, Vladimir Putin predicted:

"Artificial intelligence is the future, not only for Russia, but for all humankind. It comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world."

⁷¹¹

Impact on society

4. How can the general public best be prepared for more widespread use of artificial intelligence?

Areas of focus to enhance preparation for more widespread use of AI include, consumer level education to foster a more thorough understanding of the technology, including both the benefits and potential threats of deployment. In addition education and training for current/future developers of AI solutions will harness AI adoption more effectively and responsibly while seeking to optimize productivity through informed and efficient use of AI solutions.

Education, training, cyber security, privacy, healthcare, finance—these are just some of the areas in which AI is already making a difference, and will have an enormous impact within the next few years. The ways in which people view and experience these areas will change; social policy must change as well, and that process of change must begin sooner rather than later.

If the general public sees AI as a tool that enables and fosters humanity's growth as innovators, investigators, and idea generators, then there will be no limit to what can be accomplished. To get to that point, however, we must learn from the widespread changes that the Industrial Revolution and the advent of

⁷¹⁰ <https://www.weforum.org/agenda/2017/08/ways-to-ensure-ai-robots-create-jobs-for-all>

⁷¹¹ <https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world>

computers brought with them, and build future social policy with the lessons of the past in mind.

5. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

AI has the ability to accelerate the gap between the haves and have nots. The largest data collectors will be in an even greater position to manipulate than they currently are in more ways than they currently are. Resulting levels of disruption should not be underestimated – examples include:

- Elimination of jobs (evolution of work – some present examples suggest that initial organisation focus will be on replacing mid-level functions rather than low-level roles as this is more cost effective). This can potentially lead to an oversupply of overqualified mid-level staff in the jobs market, wage reduction and increased inequality that will be felt across geographies.
- Adaption of working practices to harness AI toolsets (retaining of workforce to use AI effectively in their respective roles) – computers will not replace everything we do but are more likely to replace specific tasks/aspects of what we do. The removal of repetitive tasks may result in an increased focus on more creative skillsets, problem solving and relationship management.
- Retraining of staff to perform new roles in new areas -> redistribution/resettlement of resources (e.g. increased use of robots in central, industrialised areas, while increasing creative and service sector roles in coastal regions – geographical impact –economic incentives to minimise this impact should be considered)

For AI, it isn't a question of a rising tide *should* lift all boats—a rising tide **must** lift all boats.

Public perception

6. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Information is power. Ensuring the public is informed as to how AI works will go a long way to removing fear of the technology. It will be important for companies and the public sector to embark and, ideally, partner on an educational campaign as AI is adopted and key changes are made. Take the banking industry, for example. As changes are made to how people bank, those changes should be accompanied with an educational campaign on why they were made, how they will benefit the customer and how the customer's information will remain secure. It will be very important for such information to accompany public sector changes as well.

Additionally, it will be beneficial for the public to understand how AI provides measurable cost savings, and increases efficiencies. Perhaps one of the most effective ways in which to demonstrate this would be to examine the healthcare implications; if AI is better at identifying early-stage cancers than people that is a powerful means of driving home the benefits of AI as a tool for good.

Industry

7. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Companies need to understand that embracing AI technology will potentially give them a competitive edge.

Proven benefits have already been seen in:

- Banking and finance – leveraging AI to address client needs and identify trends that guide financial decisions like credit scoring and worthiness; market forecasting; fraud detection; and consumer lending
- Healthcare – AI used to improve patient outcomes by making tests more effective and accurate and deriving individualized treatment plans from complex data
- Governments and militaries – implement AI to detect attacks; identify criminal or fraudulent activities and improve decision-making in emergency situations
- Law – reduce more monotonous work and human error to do key word searches in documents and free up staff to do other more higher-level thinking tasks

AI may be challenged to be helpful in areas such as social work and counseling. Marketplace sectors such as these, which require a more human, empathetic approach, will likely be sectors in which AI will play only a limited role. However, AI could perhaps be helpful with the front office paperwork to support these professions.

When exploring AI, ISACA recommends potential adopters should ask the following questions:

- What regulatory and compliance requirements must be considered as part of the AI deployments?
- What are the existing capabilities of the organization and how might implementing AI benefit or impact the organization?

- Will implementing AI require a complete overhaul of existing infrastructure? How challenging will it be to integrate AI capabilities within the existing platform and adequately govern it?
- Are the enterprise's existing resources able to support AI, or will expertise need to be recruited externally?
- How will the confidentiality, integrity and availability of large volumes of data be supported?
- Will implementation of AI affect the personnel landscape of the enterprise, and if so, how will that change be managed?

8. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

ISACA believes that well-considered, forward-thinking international norms may hold the best answers to these questions. The definitions of 'public good' and a 'well-functioning economy' will vary from nation to nation; the best solutions are those which involve compromise and consideration of how AI will evolve in the borderless, nationless world of the cognitive economy.

Corporations are not people, and they cannot be addressed in the same way individuals are. They require clear, international, consensus direction on how they will conduct business, how they will secure personal information, and how they will function in a manner that benefits humanity rather than merely enriching profits. Data is a company asset; it requires protection and management, as such. While care must be taken not to restrain innovation, care must also be exercised to ensure that corporate mega-companies' needs do not outweigh the needs of the populace.

From an AI perspective, data law will play a vital role in driving a successful management safeguard, especially given the imminent requirements of General Data Protection Regulation and the Network and Information System Directive. As organizations realign to take advantage of more data-centric, value add services, challenges with regard to data privacy, protection and security will increase in priority. Additional AI specific legal requirements will include licensing (e.g. do we have permission to process this data?) and cross border/data sovereignty challenges.

By way of a proposed approach, the implementation of rules for use of public data (e.g. EU funding competition rules) should include conditions that any data use must be in partnership with a local (UK) SME and the intellectual property must be in partnership with the SME which cannot be bought out for at least 5 years. By way of an additional example to promote investment while reducing

the risk of instability, the rules should ensure that a defined percentage of revenues and intellectual property remain in the UK for a predetermined time.

Ethics

9. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

There is an inherent ability for AI to be abused. It analyzes data at extremely fast rates in bulk amounts and often leaves questions of data ownership and access to it as fuzzy at best. Any AI needs to be carefully considered from a privacy and data ownership standpoint and should only be launched if these points are clear and legally called out for the entity controlling the AI. It is a new world and a new area of law so being as thorough as possible will be critical. Data integrity and risk of manipulation which can in turn lead to the “retraining” of AI systems to make process in accordance with an attacker’s remit is a concern. Detective controls such as periodic sampling and 4-eye process checking can provide mitigating measures. From a consumer level users will require assurance that their data is secure and is being processed in line with clearly defined and understood requirements.

Additional ethical challenges have already surfaced where AI system output has failed to take human factors into account. An automated banking service rejected mortgage applications with no knowledge of or regard to the ethnic background of the customer, leading to legal challenges

Technology industry leaders and country heads continue to push for proactive regulation of AI, especially with regard to AI enhanced weaponry systems. Elon Musk, speaking at the US National Governors Association Summer meeting in July 2017:

*"AI is the rare case where I think we need to be proactive in regulation instead of reactive. Because I think by the time we are reactive in AI regulation, it'll be too late...AI is a fundamental risk to the existence of human civilisation."*⁷¹²

10. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

It is difficult to envisage an acceptable black box AI system with the potential exception of a national security use case. Other feasible instances may arise over time.

The common perception with regard to AI execution is that you either have to be flawless or accountable. History teaches us that errors surface from black box systems over time, especially when not open to open testing; a good example being Automatic Teller Machines and bank cards which were once decreed as fully secure and were subsequently proved otherwise. Encryption protocols are

⁷¹²<https://www.theguardian.com/technology/2017/jul/17/elon-musk-regulation-ai-combat-existential-threat-tesla-spacex-ceo>

another famous case whereby open standards have strengthened their effectiveness.

The less critical the service, the more allowance may be made for a lack of AI system transparency. Where mission critical services (e.g. medical diagnosis) are in scope, these would not operate in isolation but complement a human operator who is essentially responsible for making the final decision.

In terms of AI transparency, a number of initiatives are progressing to promote an open research environment, examples include OpenAI, (a nonprofit research company founded by Tesla's Elon Musk and YCombinator's Sam Altman), while "Partnership on AI" aims to address bias, AI ethics and best practices.⁷¹³

The role of the Government

11. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

ISACA believes that it is vital that the Government be one of the key forces behind the responsible adoption, development and use of AI. Public policy measures that ensure security and privacy for data are critical. So, too, are social safety net measures, which will aid the nation's workforce in transitioning from a digital to a cognitive economy. The regulation of AI, however, must be approached carefully, so as not to stagnate innovation. In this regard, the Government's role is likely to evolve over time. It is not likely that a single measure will 'solve' the issues surrounding AI; rather, it will be an iterative process that will require constant attention, examination, and thorough, thoughtful deliberation.

The emerging consensus on approach involves a number steps: establishing governmental advisory centres of AI excellence; adapting existing regulatory frameworks to cater for AI where possible; and (perhaps) some system of registration for particular types of AI.

Learning from others

12. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

Perhaps the best lessons to be learnt come from the United States. Since 2017 began, there has been a marked shift away from technology and public policy, and this is certain to play a role in the evolution of AI, just as it is currently playing a role in the near-stagnation of cyber security in that nation. At present, the United States is providing a model of what **not** to do; avoiding discussion of a difficult issue, with repercussions that will be felt not only in corporate

⁷¹³https://thenextweb.com/artificial-intelligence/2017/04/23/artificial-intelligence-has-to-deal-with-its-transparency-problems/#.tnw_JtCJNmrI

boardrooms, but in the homes of professionals and semi-skilled workers trying to find their way forward in what is currently a digital economy, but soon to become a cognitive one.

A recent economic paper⁷¹⁴ suggests that AI and the freer flow of data could counter many of the ills that disfigured planned economies: excessive concentration of power, rent-seeking corruption, and irrational decision-making. The granular detail provided by masses of data and enhanced analysis could enable planners to offer consumers a more customized choice. Taking the basis that online platform monopolies resemble central planning institutions, the state has the potential to become a “super-monopoly” platform.

Such state-owned platforms could operate like an airport directing market-driven traffic. The airport manages capacity, sets aviation standards, balances the demands of safety, the environment and the movement of goods, and serves the needs of operators, passengers and retailers.

Ignoring the fact that AI brings with it myriad issues of concern, that all require thorough and thoughtful debate, is not the direction in which to proceed. The World Economic Forum, in contrast, has counseled that nations should immediately begin to consider the repercussions of AI-driven economies, for they will be here before we realize it. ISACA believes that the Forum’s approach is the correct one to follow; it is in this approach that forward-thinking policy can and will be developed, and that the needs of tomorrow’s workforces and marketplaces will be best met.

Apendix

A short list of some of the underlying assumptions of AI for there to be any benefit from using it, together with an associated set of challenges. For example, an underlying assumption is that AI engines will have access to vast data sets to analyse from which intelligence will be gleamed (this can be worded differently, but is only an example). The related challenges are: this could lead to unfair competition if only one or a few companies have access to certain (private) data sets; this could start an arms race in collecting the most number of data sets (including data sets which may only be obtained unlawfully or illegally; etc. etc. etc.

- There is no appreciation or understanding of “What is the optimum point at which additional data will not improve or change an intelligent decision?” the importance of this question is that not all decisions require infinite data to make the decision better. Personal data does not need to be collected beyond a certain point. E.g. a research project a few years ago worked out that on the whole only one week’s worth of personal data is required to determine

⁷¹⁴₍<http://www.jstor.org/stable/10.13169/worlrevipoliecon.8.2.0138>)

where and what any individual is doing at any time i.e. to provide a useful profile. Data collectors will likely just want to carry on collecting data even where there is no added benefit (to the consumer) or advantage to the data collector than to sell the data.

- The race for AI should mainly focus on better security models, better models for reducing possibilities for openness to abuse, faster useful results with less data (as opposed to results with more data),
- Some of the benefits around AI will be more apparent in stage 2 or 3, just like the benefits of the Cloud are being realised today even though the Cloud was talked 8-10 years ago for what it was going to do for us, but there had to be the right moment for mobile computing, apps, cameras, etc. for people to not only use the cloud but to be able to benefit using it anytime and from anywhere. Stages 2-3 (or any later stages) actually offer consumers the use of instant AI on their devices (apart from just voice apps which steal your data)
- The question of who owns the intellectual property from public data will need further clarification to attract more competitors.
- There is currently an unfair bias of AI with the biggest data collectors who have the resources to further their desire to collect more data and make use of it for commercial advantage, this is already apparent in Google and Facebook collectively responsible for the lion's share of all online advertising worldwide. Such companies only have their own research but quickly buy out anything technological advancements to maintain their position. Unless such issues are considered UK plc may end up being just a consumer of AI.
- To take advantage of this trend there needs to be focused resources on academic research, industry research and development of technical skills at all levels, together with funding to convert research into big businesses paying UK taxes.
- Any public body using AI will need to have robust models and rule sets to avoid losing public money due to hackers / scammers / criminals (home and abroad) abusing badly constructed models. To facilitate this the public sector needs to work with Risk, Governance, Assurance and Security professionals to produce guidelines for the whole of the public sector.

6 September 2017

Information Technology Industry Council (ITI) – Written evidence (AIC0176)

1 I am writing in response to the Call for Evidence made by the House of Lords Select Committee on Artificial Intelligence (AI) on 19 July 2017.

2 The Information Technology Industry Council (ITI) is the premier voice, advocate, and thought leader for the global information and communications technology (ICT) industry. Our member companies include the world’s leading innovation companies, with headquarters worldwide and value chains distributed around the globe. We advocate for policy environments that enable innovation and maximize all the benefits that ICT companies provide, including economic growth, job creation, and the tools to solve the world’s most pressing social, economic, and environmental challenges. One of the core elements of our mission, in every economy in the world, is to position our companies to be genuine partners of governments, as we believe that the interests of our industry are fundamentally aligned with those of the economies and societies in which we operate. This spirit of cooperation and partnership underlies our submission with you today.

3 As an industry, we are discussing the responsible development of AI and working to “speak with one voice,” with the hope that we can collaborate with policymakers and governments as they address AI-specific challenges and concerns. Given the longstanding close relationship between the United Kingdom and United States (where many of our companies are headquartered), as well and the structural similarities in the two economies, we are particularly interested in the direction that the U.K. takes on these important issues. Additionally, given their unique relationship, we believe the U.S. and U.K. can jointly show leadership on the international stage to ensure the responsible development of AI across the international community.

4 We have already seen see how AI benefits people and society in a wide array of fields. AI systems assist in medical diagnostics, alerting doctors to early warning signs and helping personalize patient treatments. AI powered systems can increase accessibility, fueling software programs that make digital content accessible to people with disabilities, such as helping blind and low-vision consumers “read” millions of photos, and perform auto-captioning for billions of videos. By pairing the power of AI computing with land cover maps, weather forecasts, and soil data, we can empower people with the data and tools they need to better conserve lands, improve ecosystems, and increase agricultural yields. AI-powered machines can make dangerous or difficult tasks safer for humans, opening new environments that were previously inaccessible to human exploration. And intelligent systems already monitor huge volumes of economic transactions – identifying potential fraud in real time and saving consumers

5

millions of dollars. In both the United States and the United Kingdom, startups, medium-sized companies, and larger technology companies have already developed AI systems to help solve some of society's most pressing problems. By allowing smaller businesses to do more with less, AI will jumpstart small businesses, helping them take risks and grow at faster rates than ever before.

6 As you know, AI remains an active area of research, constantly evolving and improving. While there is no single internationally recognized definition of AI, as an industry, we collectively refer to Artificial Intelligence or “AI” as a suite of technologies capable of learning, reasoning, adapting, and performing tasks in ways inspired by the human mind. And as AI evolves, we take seriously the responsibility to be a catalyst for preparing for a world driven significantly by AI, including seeking solutions to address negative, unintended consequences, and helping to train the workforce of tomorrow.

7 We respectfully suggest that the House of Lords approach each of the issues outlined in the call for evidence as opportunities for collaboration, and strive to include representatives from academia, industry, government and non-governmental organizations (NGOs) in each topic area it pursues, including: the pace of technological change, impact on society, the public perception of AI, industry responsibilities, ethical implications, and the role of government.

Industry Responsibilities and Ethical Considerations

8 While the potential benefits to people and society are promising, AI researchers, subject matter experts, and other stakeholders should and do spend a great deal of time working to ensure the responsible design and deployment of AI systems. Highly autonomous AI-systems must be designed consistent with international conventions that preserve human dignity, rights, and freedoms. As an industry, it is our responsibility to recognize the potential for use and misuse of AI technologies, the implications of such actions, and the responsibility and opportunity to take steps to avoid the reasonably predictable misuse of this technology by committing to ethics by design.

9 Technologists have a responsibility to design safe AI systems. Autonomous AI agents must treat the safety of users and third parties as a paramount concern, and AI technologies should strive to reduce risks to humans. Furthermore, the development of autonomous AI systems must have safeguards to ensure controllability of the AI system by humans, tailored to the specific context in which a particular system operates.

10 Data is a key ingredient for successful AI systems, and the availability of robust and representative data for building and improving AI and machine learning systems is of utmost importance. To promote the responsible use of data and ensure it is robust and representative at every stage of use, industry has a

responsibility to understand the parameters and characteristics of the data, to demonstrate recognition of potentially harmful bias (e.g. unfair or unintended prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair) and to test for potential bias before and throughout deployment of AI systems.

- 11 We believe tools to enable greater interpretability will play an important role in addressing transparency concerns (e.g. means to help users understand elements or factors of AI agent decision-making). Any such tools should be tailored to the unique risks presented by the specific context in which a system operates; recognizing that there is a spectrum of applications and associated risks. As an industry, we are committed to partnering with others across government, industry, academia, and civil society to find ways to mitigate bias, inequity, or other potential concerns related to AI applications.

Role of Government

- 12 We urge governments to invest in AI research and development; promote innovation; support adoption of global, industry-led, voluntary, consensus standards and best practices; partner with industry to protect personal and sensitive data; and leverage public-private partnerships (PPPs). Successful PPPs will help make AI and its deployment an attractive investment for government and industry, thereby promoting innovation, scalability, and sustainability. By leveraging PPPs, especially between government, industry and academic institutions, we can expedite AI research and development and prepare for the jobs of the future.

- 13 We encourage governments to evaluate existing policy tools and examine the applicability of existing laws and regulations before adopting new laws, regulations or taxes that may inadvertently or unnecessarily impede the responsible development and use of AI. As applications of AI technologies vary widely, over-regulating or inappropriately regulating can inadvertently reduce the number of technologies created and offered in the marketplace, particularly by smaller businesses and startups. We encourage policymakers to recognize the importance of sector-specific approaches as needed; one regulatory approach will not fit all AI applications. We encourage policymakers and regulators to work with industry to address legitimate concerns where they occur.

Impact on Society - Workforce

- 14 There is concern that AI will result in job change, job loss or worker displacement. While these concerns are reasonable, most emerging AI technologies are designed to perform specific tasks, and assist rather than replace human employees. This type of “augmented intelligence” means that portions, but most likely not all, of many employees’ jobs could be replaced or

made easier by AI. And as we saw with past productivity-enhancing technologies like electricity, the steam engine, or the microchip we stand to gain tremendously by developing and deploying this new technology. While the full impact of AI on jobs is not yet fully understood, in terms of both jobs created as well as displaced, an ability to adapt to rapid technological change is critical, and we must prepare collectively to enable our communities and citizens to do so.

15

Thank you for the opportunity provide this submission to your important work exploring AI. We hope to collaborate with the House of Lords and the British Government to ensure that AI can realize its full potential, and be used as a force for good. Please let us know if there are opportunities to assist in the future; we would welcome the opportunity to discuss any aspect in more detail.

Dean Garfield, President and CEO, Information Technology Industry Council (ITI)

6 September 2017

Innovate UK – Written evidence (AIC0220)

The Innovate UK response to the House of Lords Select Committee inquiry into what are the implications of artificial intelligence?

1. Innovate UK is the UK's innovation agency, a non-departmental public body sponsored by BEIS. It is the prime channel through which the Government incentivises innovation in business. Innovate UK is business-led. Our governing board and executive team is comprised of experienced business innovators and experts. We work with people, companies and partner organisations to find and drive the science and technology innovations that will increase productivity and exports and grow the UK economy.

2. We are working to:
 - accelerate UK economic growth by nurturing small high-growth potential firms in key market sectors, helping them to become high-growth mid-sized companies with strong productivity and export success;
 - build on innovation excellence throughout the UK, investing locally in areas of strength;
 - develop Catapult centres within a national innovation network, to provide access to cutting edge technologies, encourage inward investment and enable technical advances in existing businesses;
 - turn scientific excellence into economic impact and deliver results through innovation, in collaboration with the Research Community and Government; and,
 - evolve our funding models to explore ways to help public funding go further and work harder, while continuing to deliver impact from innovation.

3. In line with our strategy⁷¹⁵ and delivery plan⁷¹⁶, we operate across Government and advise on policies which relate to technology, innovation and

⁷¹⁵ 'Concept to Commercialisation: A strategy for business innovation, 2011-2015'.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/360620/Concept_to_Commercialisation_-_A_Strategy_for_Business_Innovation_2011-2015.pdf

⁷¹⁶ Innovate UK Delivery Plan 2016-2017:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/514838/CO300_Innovate_UK_Delivery_Plan_2016_2017_WEB.pdf

knowledge transfer. We also support Government departments to become more efficient by supporting them in developing innovative solutions through harnessing the creativity that businesses can offer.

4. Innovate UK was established in July 2007 (as the Technology Strategy Board). We have invested over £2.2 billion in innovation, and have helped more than 8,000 innovative companies in projects estimated to add up to £16 billion to the UK economy and created an average of 7 jobs per company we have worked with. Our investment over the last 8 years has meant that every £1 invested has returned up to £7.3 GVA to the economy and created 70,000 jobs. The private sector more matches that investment, doubling the power of public sector money. We work with nearly every University in the UK to stimulate the commercialisation of leading-edge academic research and innovation.
5. Driving productivity and growth is at the heart of Innovate UK's strategy and purpose and our Emerging and Enabling Technologies programme seeks to identify, and invest in, technologies and capabilities that will lead to the products, processes and services of tomorrow – those with the potential to create billion-pound industries and disrupt existing markets. Innovate UK welcomes the Committee's inquiry into The Government's approach to Artificial Intelligence. Set out below is our response to the questions raised by the Committee.

1. The current state of artificial intelligence

6. We'd like to begin by defining some terms so that our submission may be seen in the intended context.

Artificial Intelligence (AI)

- a. The Oxford English Dictionary defines artificial intelligence⁷¹⁷ (AI) as "The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages".
- b. Innovate UK considers AI to be the bringing together of science, engineering and technology with the objective of making a computing system mimic, augment or replace human activity/behaviour, across many broad and varying contexts.

⁷¹⁷ Taken from: https://en.oxforddictionaries.com/definition/artificial_intelligence

- c. Many different technologies need to be brought together to make an AI system function, including, depending upon the use case: sensing (including navigation, computer vision, 3D sensing, situational awareness etc.); natural language processing; reasoning; machine learning (ML); knowledge representation; planning; and higher-level cognition or intelligence (strong AI).
 - d. AI technologies/systems are often also called cognitive technologies/systems e.g. cognitive computing.
 - e. Artificial Intelligence can be implemented in many forms but can broadly be thought of in three progressive categories:
 - i. **Artificial “General” Intelligence**, which aims to develop a general reasoning/decision capability which can cope with diverse complex situations without external interventions and which is often thought of when people refer to AI. While great leaps forward are being made in Artificial General Intelligence, it is still very much in the domain of research and as such we believe it will be a significant amount of time before general AI is realised in any practical sense.
 - ii. **Artificial “Assistive” Intelligence**, at the other end of the spectrum, which acts as an assistant, providing insight in areas where humans have set the context, asked the questions of interest and defined the methods to be used. This type of AI is already with us and is established in practice, providing insight and personalisation in today’s services.
 - iii. **Artificial “Specialised” Intelligence**, which sits between the two. Here we are seeing emerging technologies such as machine learning, enabling knowledge based companies to make a step change in their productivity, not by replicating the skills and expertise of the human expert but by augmenting them.
7. A large proportion of current AI work focuses upon machine learning, which entails the design and development of computing systems and applications capable of learning based on their data inputs/states/outputs, without explicit programming i.e. learning by experience. Examples where machine learning systems perform better than humans in specific tasks already exist; Google’s AlphaGo computer that recently beat the Korean grandmaster Lee Sedol was based to an extent on machine learning technologies.
8. Whilst true “artificial general intelligence” technology is still considered by many as futuristic, the closely related fields of data fusion, analytics, data

mining and machine learning technologies are already finding their way in many different applications and demonstrating high value.

9. Currently most commercially used AI operates in very specific contexts. It can now outperform (better or more reliable performance, or lower cost) human operators in some of these. Examples include playing games such as jeopardy or chess, providing certain diagnoses based on medical images, identifying patterns in large data sets, rapidly resolving facial recognition etc. Such systems are limited to the specific activities and instances they are designed and implemented to support. Those solutions cannot be re-deployed to a new task without significant reworking or retraining.
10. The increasing amounts of accessible and useable data, combined with the increasing availability of affordable computing power, has helped enable significant growth and deployment in the areas of assistive and specialised AI in many areas. For example, machine-enabled translation⁷¹⁸, information search and discovery⁷¹⁹, image processing⁷²⁰, predictive text⁷²¹, automated online support services, fraud prevention, healthcare⁷²² etc.
11. Innovate UK has funded more than 260⁷²³ projects in “Artificial intelligence” with a combined value of circa £39m. Activity levels are increasing dramatically, and funding granted to AI and AI-related projects over the last three years is 65% higher than was awarded during the preceding ten years.
12. SwiftKey is a good example of what such grant investment can achieve. This company, who created a machine learning algorithm that produces more accurate predictive text, was awarded £65k of grant funding from Innovate UK in 2008/9 to help develop their predictive text App for Android and iOS devices, such as smartphones and tablets. SwiftKey recently completed a \$250m exit to Microsoft.
13. “Artificial Intelligence” companies have generated 580 successful UK fundraisings since 2007, across 301 companies, totalling £918m. There has been a significant increase in both volume and annual amounts invested since

⁷¹⁸ Google, Babylon

⁷¹⁹ Google, Bing, Yahoo et al

⁷²⁰ Magic Pony

⁷²¹ SwiftKey, Microsoft

⁷²² Mastodon C: <https://theodi.org/news/prescription-savings-worth-millions-identified-odi-incubated-company>

⁷²³ Data from RCUK Gateway: <http://gtr.rcuk.ac.uk/>

2011 with increasingly rapid growth continuing into 2017 with £334.0m being raised across 122 fundraisings during 2017 year to date (7/9/17)⁷²⁴

14. A recent report by McKinsey⁷²⁵ estimated that total spend on AI in 2016 by the global tech giants (companies such as Google and Baidu) was in the range \$20bn to \$30bn. 90% of that being spent on R&D and deployment and the remaining 10% on AI acquisitions. VC and PE financing, grants and seed investments also grew to a combined total between \$6bn and \$9bn. Machine learning, as an enabling technology, received the largest share of both internal and external investment.
15. In October 2016, the Transport Systems Catapult, funded by Innovate UK, put a self-driving autonomous vehicle on UK public streets for the first time through its LUTZ Pathfinder Project.⁷²⁶ The autonomy software running the vehicle was developed by Oxford University's Oxford Robotics Institute and integrated by Oxford University spinout, and Innovate UK funded, company Oxbotica.
16. Robotics and Artificial Intelligence (RAI) has the potential to grown national productivity, enable leaner and safer practices, enhance our quality of life and empower a more resilient society. RAI is a key strand of the government's Industrial Strategy and in Autumn 2016 it announced a national programme in RAI for extreme and challenging environments through £93m of investment, from October 2017 to March 2021⁷²⁷.

2. The pace of technological change and the development of artificial intelligence

17. AI techniques and processes have been an area of focus for research since the 1950s. Although the adoption of those early systems was relative limited, recent technology developments have led to an explosion of activity.

⁷²⁴ Data sourced from www.beauhurst.com

⁷²⁵ "Artificial Intelligence: The Next Digital Frontier":
<http://www.mckinsey.com/~media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx>

⁷²⁶ <https://ts.catapult.org.uk/current-projects/self-driving-pods/>

⁷²⁷ Phase 1 competition: <https://www.gov.uk/government/news/robotics-and-ai-apply-in-the-industrial-strategy-challenge-fund>

18. The increased availability of large scale distributed computational resource and data storage, connected by faster networking, combined with dramatic growth in data captured from processes and users, significant improvements in data processing, and a dramatic reduction in costs, has allowed the evolution of a new range of powerful AI tools.
19. These AI tools and systems are now much more widely available. Large scale computing infrastructure is easily obtainable via Cloud-based services and is being applied to the data being collected routinely in business and government in a wide range of scenarios. Many software start-ups are seizing on this opportunity to create new value in the marketplace.
20. This “Machine Learning” evolution has enabled improvements in understanding of users and systems to the point where complex systems can be created to model and predict behaviour. This has led to improvements in system performance in areas such as speech recognition, credit card fraud detection and traffic and transport service management e.g. satellite navigation tools which use historical data and real-time congestion information to inform drivers of delays.
21. The tools made available by leading industry players allow for quite sophisticated experiments to be undertaken using real data, thus building the knowledge for deploying a working artificially intelligent system. However, these tools tend to be quite specific, addressing a narrow problem, and they cannot yet be easily combined into a more generally intelligent system with predictable behaviours.
22. AI is starting to be deployed in automated image processing. and augmented and virtual reality applications. Training the ML tools remains a time-consuming process, but much innovation is taking place as part of the drive to improve security (identifying threats) and enable autonomous vehicles (recognising context).
23. There are opportunities in AI in hardware as well as in software, with a new generation of chips being developed specifically to execute computationally intensive AI algorithms. The area of Embedded AI is expanding rapidly. For example, Nvidia has been providing hardware platforms that can support the data processing for AI. UK companies such, as Graphcore who have just raised an additional \$30m, are developing specialist device architecture to support AI.

24. Further there is increasing interest in exploring the application of AI techniques in different sectors. FinTech has been an early adopter in automated trading and service delivery, but the use of AI is increasingly being explored in Agriculture, Biotech, Service Delivery and many other areas. Most, if not all, Autonomous Vehicle programmes depend upon AI techniques for success.
25. One persistent challenge is the quality and provenance of data. Many systems currently “publishing” data have been designed for a different purpose, and distributing data in real time at the right level of accuracy is often very difficult. Much of the geolocation data needs to become more accurate, especially in public infrastructure systems where such data was not previously considered important. Assuring the provenance of the digital data can necessitate ensuring that the performance of physical parts of the system, such as its sensors, are functioning properly.
26. It is our expectation that the deployment of AI and machine learning systems and tools will continue to accelerate as the barriers to entry are now lower than ever. A recent study conducted for the KTN⁷²⁸ concluded that AI is fast becoming commoditised. The tools and infrastructure required to implement it are easily accessible from the major Cloud Service Providers and the acquisition of data within business is becoming easier. It has never been easier or cheaper to deploy Artificial Intelligence tools.

3. The impact of artificial intelligence on society

27. This is a difficult question to answer. The truth is that no-one really knows, although there is a wide range of published literature examining the potential impact on business, employment, the economy and society. This indicates that AI presents both significant opportunities and real risks.
28. Different views exist on the impact of advances in robotics (physical and digital) on jobs, with both scientists and economists offering wildly varying views for how deeply automation will affect future employment. For example, in “The Rise of the Robots: Technology and the Threat of a Jobless Future” by Martin Ford, the main scenario is one where advances in robotics could wipe out jobs and deepen inequality. According to Andy Haldane, the Bank of

⁷²⁸ The study referred to, was *Artificial Intelligence: Commoditised Components for Systems* by Professor Dave Robertson at the University of Edinburgh, a long-time researcher in AI. The work was submitted to the Innovate UK ICT Industrial Advisory Board in January 2015.

England's Chief Economist, 15m jobs in the UK are "at risk of automation" by smart machines over the next two decades.

29. Artificial intelligence and robots are identified as one of the ten disruptions that could radically change the future of work in UK⁷²⁹.
30. That said, AI and the underlying technology, is expected to generate significant economic benefits, including through the creation of new jobs. A report by McKinsey⁷³⁰ estimated that the economic impact of the automation of knowledge work (through the combined advances in computing technology, machine learning, and natural user interfaces) by 2025 will be in the range \$5.2 trillion to \$6.7 trillion per year world-wide.
31. According to recent report by Tractica⁷³¹, which examines the practical application of AI within commercial enterprises, the global market for enterprise AI systems (excluding the additional associated investments or expenditures in professional services, and ICT hardware and services) will increase from \$202.5 million in 2015 to \$11.1 billion by 2024.
32. However, there is some concern that the economic and technological benefits will not be evenly distributed across society. Different scenarios are analysed in "The Future of Work - Jobs and Skills in 2030," study by the UK Commission for Employment and Skills in 2014.
33. Two recent studies by Deloitte^{732,733} concluded that UK employment is benefiting from recent technological changes and that continued success will depend on the ability of businesses, educators and government to anticipate future skills requirements and provide the right training and education.
34. Artificial intelligence, and the deployment of machine learning systems, will change the nature of work and the UK workplace and, in many areas, already has. Machines and AI-controlled systems don't perform *jobs*, they automate *tasks*, both physical and intellectual. They change the skills required to

⁷²⁹ Similarities exist with the related analysis for the USA, in "The Future of Work," MIT Technology Review, Dec. 2015.

⁷³⁰ McKinsey Global Institute "Disruptive technologies: Advances that will transform life, business and the global economy," 2013.

⁷³¹ "Artificial Intelligence for Enterprise Applications," 2015; it examines the practical application of AI within commercial enterprises.

⁷³² "London futures Agiletown - the relentless march of technology and London's response," Deloitte, 2014

⁷³³ "From brawns to brain- the impact of technology on jobs in the UK," Deloitte, 2015

perform the job and the nature of the work itself. Notable examples include stock control, automated trading, e-retail, satellite navigation systems, etc.

35. Access to AI and AI-enabled systems will also benefit the UK population in a number of other areas. For example, machine learning systems which help identify and prevent credit card and other financial fraud, navigation systems and routing algorithms which will help avoid traffic congestion and minimise travel times, improved recommendation engines to help the search and discovery of creative content.

36. The use of AI and AI-enabled systems will also allow more affordable and more accessible access to services. For example, more rapid and accurate healthcare diagnosis, improvements to the justice system through increased efficiency and accessibility, optimised insurance services focused on the genuine needs and requirements of the customer. All of which have wider societal benefits.

37. The legal implications of artificial intelligence are under consideration but will need ongoing and continuous review as the technology and the capability evolves over time. That evaluation will need to consider not only the technology but also the potential impacts and changes to human behaviour when interacting with an AI-enabled system⁷³⁴.

4. The public perception of artificial intelligence

38. Innovate UK has not conducted any research in this specific area but we would make the following points in relation to business-led innovation.

39. In order for new products, services or processes to be sustainable over the longer term they must be acceptable to the public and society. It is in their own, as well as society's, best interests that businesses innovate responsibly. In the field of AI, we have observed both exaggerated fears and the thoughtful raising of reasonable concerns.

40. Some science fiction has raised concerns in the mind of the public on the safety of "AI" systems.

41. There are some public concerns that driverless cars might not respond appropriately to the real events that happen on roads, such as interaction

⁷³⁴ <https://www.lawsociety.org.uk/news/blog/do-we-need-new-law-or-legal-concepts-to-govern-ai-and-machine-learning/>

with cyclists, pedestrians and other drivers. Concerns around autonomous vehicles, combined with the fear of potential loss of employment, require some thoughtful public engagement to be conducted.

42. While autonomous systems are potentially safer, they are unlikely to be perfect. AI controlled systems will make mistakes, just as humans do. However overall, autonomous vehicles will be significantly safer than cars controlled by humans, who are also known to make crucial mistakes when faced with an impending accident⁷³⁵.
43. To date, most people's experience of an autonomous AI-enabled system is limited to the 'recommendations' features in online retail sites such as Amazon, voice recognition systems and chatbots. Even relatively well-developed AI-enabled assistants such as Siri and Alexa, and those installed in modern vehicles, are voice-controlled and have exposed the public to the vagaries of interaction through natural language, with occasional misunderstandings and failed comprehension.
44. These concerns have been writ large in the Tay experiment undertaken by Microsoft⁷³⁶, a very short-lived attempt to train an AI system using real world social media behaviours. A Twitter 'chatbot' was launched in March 2016 and which was designed to "engage and entertain people where they connect with each other online through casual and playful conversation". However, the agent was shut after only 16 hours after it started to post inflammatory and offensive messages. Tay was merely reflecting and building on the type of content it was being exposed to, but this is an example of what can happen when these kind of automated learning systems are 'released into the wild' without sufficient controls and oversight.
45. The Future of Life organisation has published a balanced description of AI and its short and longer-term goals⁷³⁷, together with a sensible reduction of existing "public" concerns regarding AI safety. It reports the concerns raised by leaders in the scientific and technology community, publicised earlier this year when comments were made by Professor Stephen Hawking of Cambridge University.

⁷³⁵ Ref: <http://www.autonews.com/article/20160110/OEM06/301119963/autonomous-vehicles-will-be-safer-not-perfect>

⁷³⁶ Ref: [https://en.wikipedia.org/wiki/Tay_\(bot\)](https://en.wikipedia.org/wiki/Tay_(bot))

⁷³⁷ See <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/>

46. Another concern which is widely discussed in the media is the impact on work and working opportunities, with studies beginning to report the impact on professional roles that AI will replace. The Economist, Fast Company, Forbes and Financial Times have all published stories on this topic in the past year.⁷³⁸

5. The sectors most, and least likely, to benefit from artificial intelligence

47. Nearly every technology sector in the UK has the potential to be affected by, and potentially to benefit from, the use of AI's digital automation and decision-making capabilities.

48. It is too early to say definitively where the impacts of AI will be the greatest. In the same way that robots gained traction in those parts of the economy where repetitive or manual tasks are involved, AI will probably have the most impact in sectors where cognitive or complex tasks dominate and which can in turn be automated. As robotics affected mostly 'blue collar' workers, so AI might affect 'white collar' businesses⁷³⁹.

49. A report from PWC suggests that up to 30% of jobs in the UK could potentially be at risk from automation by the early 2030's⁷⁴⁰, with the highest risk being in sectors such as transportation and storage; manufacturing; wholesale and retail.

50. The application of AI tools in service delivery, is now on the threshold of enabling the automation of call handling and response to service requests. This could replace many thousands of service jobs in the UK and offshore. A key requirement is the ability to convince a caller that the service is intelligent, a combination of improved voice recognition, vocabulary and accent capabilities and the ability to respond as expected to a request.

⁷³⁸ <https://www.fastcompany.com/3066620/this-is-how-ai-will-change-your-work-in-2017>
<https://www.ft.com/content/f809870c-26a1-11e7-8691-d5f7e0cd0a16>
<https://www.forbes.com/forbes/welcome/?toURL=https://www.forbes.com/sites/jeannemeister/2017/03/01/the-future-of-work-the-intersection-of-artificial-intelligence-and-human-resources/&refURL=https://www.google.co.uk/&referrer=https://www.google.co.uk/>
<https://learnmore.economist.com/story/57ad9e19c55e9f1a609c6bb4>

⁷³⁹ See also: Jerry Kaplan "Humans Need Not Apply: A Guide to Wealth and Work in the age of Artificial intelligence": <https://www.youtube.com/watch?v=hoDxcO2EOHM>

⁷⁴⁰ Taken from 'Will robots steal our jobs? The potential impact of automation on the UK and other major economies': <http://www.pwc.co.uk/economic-services/ukey/pwcukey-section-4-automation-march-2017-v2.pdf>

51. The development and deployment of AI-controlled autonomous vehicles continues apace and the use of autonomous systems is already not limited to just cars. Tesla, Uber, Google, Volvo and others are developing autonomous trucks⁷⁴¹, while companies like Rolls Royce are developing both remotely controlled⁷⁴² and automated vessels for commercial shipping⁷⁴³.
52. The promise of AI to improve agriculture and healthcare is also attracting attention, given the potential to better optimise crop yield, insecticide and fertiliser usage, and better diagnose and treat patients.
53. Future Infrastructure Systems will depend upon the application of AI to respond to events and situations and orchestrate a suitable change in demand or usage. Examples will include the use of electricity for charging electric vehicles (minimising the draw on the distribution network); the assembly and management of a convoy of vehicles operating autonomously on a crowded motorway; the intelligent end-to-end planning of flights from gate to gate, minimising the usage of critical paths such as taxiways and runways and air traffic slots.
54. There are also potential applications in professional services, including in consultancy and the law. Innovation in services and service delivery will become a key differentiating factor⁷⁴⁴. Businesses working in knowledge-based services therefore have a unique opportunity to use this augmentation of human intelligence with Specialised Artificial Intelligence to create a broader range of solutions to the problems they are working on. This in turn creates an increased probability of finding better outcomes for their clients and therefore stronger positions in the global market.
55. Despite recent progress in specific areas, much technology development remains to be done and there are other issues to be resolved before truly widespread adoption can take place.

⁷⁴¹ "Here's how Tesla, Uber, and Google are trying to revolutionize the trucking industry": <http://uk.businessinsider.com/autonomous-trucks-tesla-uber-google-2017-6/>

⁷⁴² Ref: <https://www.rolls-royce.com/media/press-releases/yr-2017/20-06-2017-rr-demonstrates-worlds-first-remotely-operated-commercial-vessel.aspx>

⁷⁴³ Ref: <https://www.rolls-royce.com/~media/Files/R/Rolls-Royce/documents/customers/marine/ship-intel-160316.pdf>

⁷⁴⁴ The Future of Legal Services: <https://www.lawsociety.org.uk/news/stories/future-of-legal-services/>

56. Although many jobs are likely to be threatened by new AI and Autonomous Systems technologies⁷⁴⁵, history (for example the agricultural and industrial revolutions) tells us that new technologies also create entirely new forms of employment that simply did not exist before. Overall, technology has created more jobs than it has destroyed in the last 144 years⁷⁴⁶.

6. The data-based monopolies of some large corporations

57. Many machine learning systems are based to a large degree on algorithms that 'learn' their function by exploring the data sets they are presented with. In some instances, the algorithms are trained on subsets of the data, and the validity of the learned algorithm validated by testing on the rest of the 'unseen' data.

58. To a significant degree therefore, the ability to develop ML-AI systems depends upon having access to the relevant data. This has led a number of large corporates to accumulate vast volumes of data, of various kinds, in order to unlock the potential value in the AI systems it will enable them to develop.

59. Not all such data is 'personal' or 'user' data. Often data sets might relate to the physical parameters of processes or systems, such as the functioning of an engine, the condition of a piece of factory machinery, or the state of the weather. ML has the potential to derive great value from such data, and UK companies should be encouraged to explore what benefits they might accrue in areas such productivity improvement, product quality, process optimisation and reduction in working capital.

60. Despite its potential value, the accumulation of this type of data set goes largely unnoticed. It is 'personal' data that has generated the most media attention.

61. The leading online digital media service providers, e.g. Google, Amazon, Netflix; large retailers; e.g. Tesco, Sainsbury's, Boots; digital advertisers and credit card companies have become increasingly adept at this, accumulating

⁷⁴⁵ Ref: http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

⁷⁴⁶ Report by Deloitte: <https://www2.deloitte.com/uk/en/pages/finance/articles/technology-and-people.html>

large volumes of data related to the user and their services⁷⁴⁷. In the case of Google, this includes the aggregation of data related to place and location in the world. This allows them to offer tailored services to users and potentially to offer distribution and logistics services to the industry.

62. Google's investment in the British company DeepMind in 2014 increased their investment, and capability, in machine learning and neural networks, a particularly important development when combined with Google's substantial mountain of owned data.

63. Google and other organisations also benefited from access to many disparate, significantly large, data sets, enabling access to a pool of data that is unprecedented and which creates significant commercial opportunities. Such opportunities are inaccessible for businesses without that same quantity or quality of data.

64. All major companies are now in a position where they "own" user data that may be analysed and modelled for the improvement of their business. In the case of personal data, the incoming General Data Protection Regulation will bring in important new measures to improve transparency of these processes and allow users to review the information being maintained and request its removal, where required.

7. The ethical implications of artificial intelligence

65. As the use of robots and autonomous systems grows, and more research is carried out in this area, it is important that ethical issues associated with their use are considered. In April 2017, BSI published 'BS 8611 Ethics design and application robots'⁷⁴⁸.

66. Artificial Intelligence Systems that develop understanding through analysis of large volumes of data, and/or use networking tools to model patterns of behaviour or phenomena, have the potential to evolve beyond the expected limits envisaged by their programmers and operators. The issue then becomes one of behaviour control and how such evolving systems will be arbitrated.

⁷⁴⁷ The Economist, "Getting to know you": <https://www.economist.com/news/special-report/21615871-everything-people-do-online-avidly-followed-advertisers-and-third-party>

⁷⁴⁸ See: <https://www.bsigroup.com/en-GB/about-bsi/media-centre/press-releases/2016/april/Standard--highlighting-the-ethical-hazards-of-robots-is-published/>

67. There are concerns in some quarters about the amount of data being acquired and stored by various technology companies and service providers, and their subsequent use or sharing of it. The individual to whom the data relates is unlikely to have much awareness of their current digital footprint or the extent to which their data is being shared and used.
68. The use of personal data can often benefit the customer. However, the extensive accumulation of data from a wide range of different online sources make it increasingly possible to identify people, their homes, family members, employers, vehicles, phones, credit cards, medical records, financial information etc. Issues will arise if this information is used to infringe individual privacy or adversely affect the availability or cost of services they rely up. Care must be taken to ensure that such information is securely protected against unwarranted access and when “personal” information is required only the minimum information is offered.
69. Machine learning algorithms will pick up on any bias in the data they are given to learn from, conscious or otherwise, and it is very easy for bias to be unwittingly introduced into those algorithms by their designers⁷⁴⁹. This can cause significant concern when those machine learning systems subsequently interact with the public⁷⁵⁰.
70. In May 2016 a report by ProPublica⁷⁵¹ claimed that a computer programme used by a US court was biased against black prisoners. While the company that supplied the software, Northpointe, disputed the conclusions of the report⁷⁵² this example reflects the public fears of autonomous systems and does raise concerns about the deployment and transparency of such systems.
71. It is still relatively early for standards to be applied in this field, the tools and their practice in AI applications are still largely handcrafted and depend upon experts in several areas to implement correctly.

⁷⁴⁹ Ref: <https://theconversation.com/growing-role-of-artificial-intelligence-in-our-lives-is-too-important-to-leave-to-men-82708>

⁷⁵⁰ See <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>

⁷⁵¹ ProPublica report: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

⁷⁵² Northpointe response: <https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html>

72. For these reasons, it is our expectation that, at least in the short to medium term, future commercial applications will retain a human “in the loop” to oversee, guide or improve the use of such systems.

8. The role of the Government

73. The role of government in relation to the development of AI and machine learning is to help ensure that future developments in AI benefit the UK; both economically through supporting businesses; and societally through ensuring mitigation of the risks. In short, the Government has an important role in helping to prepare for the future.

74. System resiliency, cyber-hardening, cyber-security: concern over the vulnerability to hacking of Artificial Intelligence Systems should be taken extremely seriously and systems to prevent unauthorised access or modification put in place. It is vital that a high level of trust is maintained in the use of these systems and the legal aspects of responsibility and liabilities need to be defined.

75. Public awareness and acceptance of the actual risks of AI and autonomous systems needs careful consideration and active intervention. The public is only partly sighted on the opportunities and issues associated with AI. If the wider benefits are to be realised, both legitimate concerns and baseless fears must be addressed.

76. In March 2016 Sciencewise facilitated an event⁷⁵³ to map issues of potential concern to the public in the area of ‘Robotics and Autonomous Systems’. 73 people attended the event including Government policy makers, academics and industry leaders but given the most recent advances in the AI sector, further investigation and public engagement activities should be undertaken.

77. Allocating responsibility, blame and costs/liabilities should things go wrong are issues yet to be resolved for AI controlled autonomous or automated systems. The government has a role to play in helping shape that dialogue and helping to resolve the societal and legal concerns.

78. Another very important role for the government is to help support UK companies to derive maximum benefit from this technology; both on the

⁷⁵³ Sciencewise meeting notes: <http://www.sciencewise-erc.org.uk/cms/assets/Uploads/Meetings/RAS-Meeting-Notes-7-March-2016-FINAL.pdf>

supply side (machine learning focused businesses) and the end-user organisations, whose core business activities might not be in the AI domain per se but who could derive great value from it.

79. Government has a very important role in the provision of education, training and skills; in schools, in the research base, and at the interface to business. This is the case for the new roles that these technologies will create, as well as for the upskilling of many existing technical roles that will utilise these technologies in the future. The earning power of those with these new skills would be significantly higher. KTP has a lot to offer here, in both cases, and should continue to be supported.
80. There are also opportunities for Government to use AI to improve its delivery of public services, and optimisation the use of national infrastructure.
81. The Digital Catapult, funded by Innovate UK, is working to remove barriers around access to large quantities of high quality training data for machine learning models; helping companies with access to the computational power needed to extract value from data; supporting the adoption of AI and machine learning technology by companies of all sizes; and continuing to help convene new commercial relationships.⁷⁵⁴.
82. Opportunities exist within the field of AI to support experimentation and learning and Innovate UK, the Digital Catapult Centre and Turing Institute will continue to act as areas of focus in this area.

9. The work of other countries or international organisations

83. All countries in the developed world are exploring how to use AI and AI-based systems, and many have initiated studies or investigations into the implications and potential benefits of this technology. For example, the UK is undertaking a major AI review led by Dame Wendy Hall and Jérôme Pesenti⁷⁵⁵ and in May of this year the French Parliamentary Office for Scientific and Technological Assessment (OPECST) released its findings⁷⁵⁶.

⁷⁵⁴ <https://www.digitalcatapultcentre.org.uk/>

⁷⁵⁵ "Digital Strategy to make Britain the best place in the world to start and grow a digital business": <https://www.gov.uk/government/news/digital-strategy-to-make-britain-the-best-place-in-the-world-to-start-and-grow-a-digital-business>

⁷⁵⁶ "Toward a Controlled, Useful and Demystified Artificial Intelligence": <https://www.senat.fr/rap/r16-464/r16-464-syn-en.pdf>

84. The USA is a global leader in the application of AI through numerous companies that are developing and adopting it: Google, Amazon, Facebook, Microsoft, Netflix and the like. IBM has made a major investment in a machine learning capability in Watson and is seeking to acquire new insight, and customers, through analysing large volumes of data in different industry areas.
85. China will be a major player in this area and Germany has the engineering discipline to implement these systems. Both countries, and Japan, are leaders in the application of industrial robotics where the natural evolution is towards increased intelligence and autonomy.
86. The UK economy is significantly dependent on the service sector and the service industry will be enormously affected by advances in AI and AI-enabled systems. The UK simply must not be left behind in this race for innovation as the potential impact on the UK is as significant as for any other developed nation.

Evidence submitted on behalf of the Innovate UK by:

Dr Ruth McKernan, CBE
Chief Executive, Innovate UK

13 September 2017

The Institute of Chartered Accountants in England and Wales – Written evidence (AIC0041)

What are the implications of artificial intelligence?

ICAEW welcomes the opportunity to comment on the call for evidence What are the implications of artificial intelligence? published by House of Lords Select Committee on Artificial Intelligence on 19 July 2017, a copy of which is available from this link

This response of 1 September 2017 has been prepared on behalf of ICAEW by the IT Faculty. Recognised internationally for its thought leadership, the Faculty is responsible for ICAEW policy on issues relating to technology and the digital economy. The Faculty draws on expertise from the accountancy profession, the technology industry and other interested parties to respond to consultations from governments and international bodies.

ICAEW is a world-leading professional accountancy body. We operate under a Royal Charter, working in the public interest. ICAEW's regulation of its members, in particular its responsibilities in respect of auditors, is overseen by the UK Financial Reporting Council. We provide leadership and practical support to over 147,000 member chartered accountants in more than 160 countries, working with governments, regulators and industry in order to ensure that the highest standards are maintained.

ICAEW members operate across a wide range of areas in business, practice and the public sector. They provide financial expertise and guidance based on the highest professional, technical and ethical standards. They are trained to provide clarity and apply rigour, and so help create long-term sustainable economic value.

Copyright © ICAEW 2017
All rights reserved.

This document may be reproduced without specific permission, in whole or part, free of charge and in any format or medium, subject to the conditions that:

- it is appropriately attributed, replicated accurately and is not used in a misleading context;
- the source of the extract or document is acknowledged and the title and ICAEW reference number are quoted.

Where third-party copyright material has been identified application for permission must be made to the copyright holder.

MAJOR POINTS

1. We believe that machine learning (our focus in artificial intelligence, or 'AI') is potentially a very powerful tool for society. We emphasise three particular capabilities of machine learning in this regard – learning from enormous amounts of data, building complex and changing patterns from data and achieving high levels of consistency. These powers can 'turbo charge' human capabilities and enable significantly better decisions. Whether AI will deliver on its promises in practice, though, remains to be seen, as it is still in early stages of mainstream adoption.
2. The accountancy sector will be able to deliver more value to organisations and the economy because of AI. AI-based systems and tools can potentially improve the efficiency of accountants, enable them to focus on areas of highest business risk, and improve the quality of business decisions. AI may also enable them to measure and analyse a wider range of business activities, and improve other areas of decision making and accountability.
3. For many people, the most direct impact of AI will be on their jobs. We take a positive view of the future and believe that humans will continue to find many ways to contribute to economies and societies, alongside machines. However, there will be significant changes to the business and employment environments. This will change the skills that younger generations need and the education system needs to recognise that. Life-long learning will also become vital, and our training infrastructures must be updated to cope with this shift.
4. We do not believe that it is possible to regulate a technology such as 'AI' in isolation. There is no common definition of what AI is. Furthermore, it increasingly just permeates across many activities in our lives. However, existing regulators urgently need to consider the impact of AI-based systems (now and in the future) on their sector. They need to encourage investment where it will deliver improvements to the sector, and decide how to manage the risks.

RESPONSES TO SPECIFIC QUESTIONS

Q1: What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

5. While Artificial Intelligence (AI) is a broad field, we primarily focus on machine learning techniques. We characterise the current state of machine learning in the following way:
 - a) There have been great improvements in the accuracy of machine learning in recent years, primarily due to two factors - high volumes of data and greatly improved processing power. The combination of these factors has led to far higher levels of accuracy in the predictions made by machine learning models, making them more useable in the real world.

- b) Mainstream business adoption is still in early stages. While companies in the internet sector, and some parts of financial services, have been developing and using these techniques for some years, we must not lose sight of the reality of most businesses, who are a long way behind in their adoption of many technology trends, including AI.
6. When looking to the future, two different aspects need to be considered. Technology capabilities will of course be important, and there will be many technical challenges to overcome, such as continuing improvements to processing power, and the ability of computers to cope with even greater volumes of data.
7. However, actual adoption and use will be driven by a wide variety of factors, including:
- a) Practical application – for example, identifying the most relevant use cases, having appropriate software on the market, having enough good quality data for accurate results and changing processes to maximise the benefits.
 - b) Economics – building business cases for the development and adoption of AI systems.
 - c) People – in particular, ensuring trust in the systems and their outputs, and having enough skilled people to develop, implement and use systems.
8. We broadly see two models of adoption – conscious adoption of AI systems to solve specific business problems, which will require significant resources and skill; and unconscious adoption of AI, whereby AI capabilities are simply integrated into existing business software (primarily based in the cloud). These different models will help smaller businesses to benefit from AI capabilities, without needing the skills and resources of larger businesses. However, we do see great uncertainty about the timeline for change and the extent of adoption.

Q2: Is the current level of excitement which surrounds artificial intelligence warranted?

9. We believe that machine learning is potentially a very powerful tool for society, and emphasise three particular capabilities in this regard.
- a) Machine learning systems can process enormous amounts of data, far beyond human capabilities, offering opportunities to develop much better learning and knowledge. AI, in this context, becomes an essential tool to help us make sense of all the data being generated today and in the future.
 - b) Machines can learn far more complex and changing patterns than we can and therefore be far more effective in environments that we see as unpredictable.
 - c) The consistency of decision making by algorithm can also improve the quality of decision making and take out many human biases (although

we note the risks of models perpetuating systemic bias based on historic data).

10. These powers can 'turbo charge' human capabilities and enable better, quicker and more consistent decisions. Whether AI will deliver on its promises in practice, though, remains to be seen. As stated earlier, mainstream business use of AI is still in early stages, and there are many open questions about its real world effectiveness across different domains.

Q3 How can the general public best be prepared for more widespread use of artificial intelligence?

11. For many people, the most direct impact of AI will be on their jobs. We take a positive view of the future and believe that humans will continue to find many ways to contribute to economies and societies, alongside machines. However, there will be significant change in the business and employment environments. While there is nothing new about technology impacting jobs, we must not downplay the level of disruption likely to be experienced in many individuals' jobs and lives, and policy makers should be thinking about how to mitigate the effects of this. While we do not necessarily subscribe to solutions such as Universal Basic Income, there is an urgent need for greater debate between policy makers, and across society as a whole, of the potential impact of AI on jobs and mitigating actions.
12. It is very likely that AI will lead to significant changes in the skills demanded by employers and this will require rethinking at two levels:
 - a) Younger generations will need to build the right skills to work in this changing environment. Being able to work effectively with technical specialists, and doing a certain amount of technical and analytical work, will become central to more business jobs. There will also be greater emphasis on the uniquely human skills that will complement computers, such as empathy, story-telling, persuasion, critical thinking and creativity.
 - b) There will need to be greater focus on reskilling and retraining throughout our lives. Technology will continue to change very fast and humans will need high levels of adaptability and resilience, as well as the acquisition of new skills, to keep up. Life-long learning will become vital, and training infrastructures must be updated to cope with this change.
13. We are already experiencing significant change in accountancy jobs and skills because of technology, and AI will amplify those. We see, for example, reduced transactional accounting work and greater emphasis on gaining and applying new insights from data. This typically needs more skills in data, and we are continually updating our qualifications to incorporate more technology and data skills in response to market demands. We also emphasise personal and professional skills such as critical thinking and communication.

Q4 Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

14. No comment

Q5 Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so?

15. No comment

Q6 What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

16. We focus our comments on the accountancy sector, based on our recent report Artificial intelligence and the future of accountancy. We believe that the accountancy sector will be able to deliver more value to organisations and the economy as a result of AI. The profession supports and improves business and investment decision making through processing, organising, analysing and communicating information. Therefore, machine learning should be a powerful tool for the profession.
17. Use of machine learning in accountancy is still in early stages and builds on existing capabilities around big data and data analytics. Some large firms and finance functions are investing heavily in these new technologies. Smaller firms and businesses are generally some way behind. Examples of the kind of use cases being discussed include:
- a) using machine learning to model 'normal' transactions and therefore identify 'abnormal' transactions more easily in forensic accounting and audit – this can focus resources on areas of greatest risk
 - b) using deep learning capabilities around text to analyse contracts and identify specific risks or liability – this can increase efficiency and enable more analysis of text-based documents
 - c) improving financial forecasting and planning through machine learning models - this can improve business decision making
 - d) using machine learning to automatically code account entries in accounting systems - this can free up the time of accountants to focus on more value-adding, advisory work
18. These examples therefore cross all aspects of the profession, have relevance to organisations of all sizes, and provide a range of potential benefits.
19. In the longer run, AI, combined with many new sources of data, gives accountants the opportunity to use their skills in new areas and contribute more to the economy. Accountants currently only measure and analyse a small subset of business activities, due to lack of data in areas such as intangibles. Improvements in data and technology provide opportunities for the profession to measure, analyse and improve decisions in many other areas, for example activity related to the UN Global Sustainability Goals.

Q7 How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

20. No comment

Q8 What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

21. AI is a decision-orientated technology. It produces predictions that can be used to inform and automate decision-making processes. This, therefore, raises many ethical questions about how decisions are made, who makes them and how to ensure accountability, including:
- a) Are models producing the expected results, and complying with principles such as fairness and privacy?
 - b) Who is responsible and accountable for decisions made by the models, and how can errors be corrected?
 - c) What is the relationship between models and human judgement, and to what extent can humans override AI systems?
22. There are many documented examples of bias and discrimination in outputs from big data models, or algorithms being relied upon inappropriately, due to factors such as poor understanding of the data being used or a lack of feedback mechanisms. These impacts will be amplified by machine learning and therefore ethics should be emphasised as an integral part of developing and using AI systems.
23. We welcome the efforts within the data science community to develop thinking and frameworks around ethics. The development of models involves many choices about data and algorithms that have ethical dimensions. Therefore embedding ethical thinking into the model-building process is vital.
24. However, the ethical dimension also needs broader discussion. While long-standing principles are unlikely to change, there may need to be fresh thinking about new scenarios or questions raised by AI. Technology can transform ethical dilemmas from theoretical discussions into real world problems. AI, and big data more broadly, has the potential to do this in many areas, from the decisions of autonomous vehicles to the personalisation of insurance coverage.

Q9 In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

25. No comment

Q10 What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how

26. We do not believe that it is possible to regulate a technology such as 'AI' in isolation. There is no common definition of what AI is. Furthermore, it increasingly just permeates across many activities in our lives. However, established regulators urgently need to consider the impact of AI-based systems (now and in the future) on their sector. They need to encourage investment where it will deliver improvements to the sector, and decide how to manage the risks.
27. The risks fall into many categories. They include security and resilience of systems, accuracy and assurance over the outputs, systemic risks where systems are interacting, concerns of consumers around personal data and how to redress wrongs created by AI systems. Clearly, these will vary across industries and regulators need to build their own understanding of AI in their context.
28. This is not an easy task, as most regulators will lack technical skills in AI and will have to build up their knowledge of the topic, as well as its specific application to their sector. Professional and industry bodies such as ICAEW can play an important role in supporting regulators, gathering insights into how technologies are being used in practice, and providing a more consistent view of experience across the sector. We are keen to contribute further in this area.
29. There are also significant challenges around pace of change. Regulation by its nature is reactive and slow-moving. But we are likely to see a fast pace of change in many sectors and regulators will need to develop more proactive approaches to cope with that, engaging early with innovators to identify issues. The Financial Conduct Authority's 'sandbox' approach to innovation in Fintech has been broadly recognised as one of the reasons behind the UK's success in this field. There may be lessons to learn for other regulators on how to manage early engagement and constructive dialogue with innovators.
30. Regulators should also be investigating ways of using AI themselves to improve their regulatory activities. Many regulators are overwhelmed by greater volumes of data, and AI can provide insights from it as well as enable better predictive capabilities. This can allow better targeting of their resources, as well as earlier identification of issues.
31. More generally, the government should focus on the skills agenda, as highlighted earlier. This includes ensuring the number of specialists in AI grows quickly to meet market demand, ensuring that younger people more broadly are learning the skills they need to operate in a world full of AI, and supporting adults to reskill for the changing business environment.
32. The government can also invest in AI capabilities in public services to improve how they are done. Areas such as healthcare and transport present tremendous opportunities for AI to cope with huge amounts of data and support better decision making at all levels. The government should aim to be an exemplar in these areas, to improve public services, encourage others

to adopt the technology and actively support the development of the AI industry in the UK.

Q10 What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

33. No comment

1 September 2017

Institute of Mathematics and its Applications – Written evidence (AIC0107)

Written evidence to the inquiry of the House of Lords Select Committee on Artificial Intelligence, submitted by the Institute of Mathematics and its Applications.

1. The Institute of Mathematics and its Applications (IMA) exists to support the advancement of mathematical knowledge and its applications and to promote and enhance mathematical culture in the United Kingdom and elsewhere, for the public good.
2. It is the professional and learned society for qualified and practising mathematicians, with a membership of around 5,000 comprising of mathematicians from sectors including research, education, industry, commerce and the public sector, as well as those with an interest in mathematics.
3. This response addresses questions 1, 4,6,8,9 and 11 of the inquiry.
4. **Question 1:** *What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?*
5. **Answer to Question 1 (paragraphs 5-12):** “Artificial Intelligence” is a very broad church. There is a fundamental distinction between Specific (or Weak) AI, that intends to address specific questions, and General (or Strong) AI, which aims to generate new questions and generally mimic human consciousness. This IMA submission concerns itself with Specific AI only. Two very different areas within Specific AI have seen enormous growth in the past ten (or more) years, and the trend shows no sign of stopping. We confidently expect the progress over the next 5 and 10 years to be at the same rate, by which time these areas will be almost unrecognisable from where they are now. There is no reason to expect progress to stop there, but it’s hard to envisage what future progress will be in any detail. One open question is whether there will be any convergence between the areas in the direction of a reasoned explanation of machine learning (what the few researchers in this area are calling “Explainable AI”).
6. **Machine Learning.** This is essentially the automatic detection of patterns in data (a process often referred to as “training”), and the use of these patterns to make decisions about new data. This sub-area has grown tremendously, and very visibly, over the last ten years, to the point where, in the public eye, it *is* artificial intelligence, and has practically taken over the meaning of some

traditional words, to the point where the last House of Common Science and Technology Committee launched an inquiry into “algorithms”, meaning “machine learning”. Our evidence to that Committee (as updated) is at <https://ima.org.uk/6910/the-debate-about-algorithms/> .

7. A significant development in the short period since that call has been a growing realisation that machine learning is very susceptible to “hacking”, in two essentially different ways.
 - 7.1 A given machine learning algorithm, say for image recognition, can be tricked into recognising an image for something that it isn’t. This has been recently researched in Evtimov *et al.*, 2017 (see *References*, paragraph 24.2), who trained a typical recogniser on a standard database of U.S. road signs, and then presented this recogniser with subtly modified road signs, one example of which is the image below, which was recognised as a 45 mph speed limit, rather than a STOP sign. As well as the specific point that this study is making, this also illustrates the fragility of much machine learning: the humans may think the machine has learned to read “STOP”, but that doesn’t seem to be what’s happening.
 - 7.2 The learning process itself can be subverted, to produce an “algorithm” which works on the training and testing data, but behaves differently on data containing certain pre-defined triggers. Gu *et al.*, 2017 (see *References*, paragraph 24.4) studied this, again on U.S. road signs, and they were able to use inconspicuous markers to trigger misrecognition. This is all the more worrying as people tend to outsource the training of machine learning “algorithms” to cloud providers, and do not necessarily have as much control over the integrity of the process
 - 7.3 Related to that, Gu *et al.*, 2017 (see *References*, paragraph 24.4) also point out that many developers of neural net machine learning “algorithms” use, and indeed are encouraged to use, “transfer learning”, where a neural net trained on a related task is used as the basis for the new network, which is merely retrained, far more cheaply, on the new training data. They demonstrated that their subverted version of the U.S. road sign recogniser retained the hidden property of misrecognition on certain triggers even after being retrained on Swedish road signs.



Image: STOP sign that Machine Learning, trained on a standard database of road signs, recognises as a 45MPH speed limit. [Evtimov et al.,2017, (see *References*, paragraph 24.2)].

8. The growth of machine learning has been caused by four factors.

8.1 Better algorithms for doing the pattern detection, notably the family known as “deep learning” (where “deep” refers to the number of layers of neural net, and has no connection with the intellectual depth of the patterns learned).

8.2 Better (and cheaper) hardware, not just traditional computers, but the “General Purpose Graphical Processing Units” (GPUs), which on central machine learning tasks can be 10-100 times faster than a conventional computer. Indeed, Google has commissioned its own variant of this, the Tensor Processing Unit, which is a further factor of 10 faster. [<https://drive.google.com/file/d/0Bx4hafXDDq2EMzRNcy1vSUxtcEk/view>]

8.3 More “easy-to-use” software and bite-sized training, so that many people can claim themselves to be “data scientists” with only a vague understanding of the theory, and alas no exposure to the ethics.

8.4 More Data. The “Big Data” effect, and the fact that many more data are publicly available, has led to both technology push, the ability of the “data scientist” to make more predictions, and the business pull to make more predictions. Whereas advertising rights in major sporting events are still

bid for by humans, the right to show a 17-year old youth an advertisement before a YouTube video is bid for, and auctioned off, by computers following machine learning algorithms.

9. **Automated Reasoning.** This is the ability of computer programs to make formal logical deductions. Unseen by the public, and indeed by many professionals, this area has made great practical advances in the last twenty years. It occasionally surfaces in the popular media as a new mathematical result is formally proved by computers, but in fact has made tremendous inroads into daily life without being noticed. Its practical deployment was started in the field of computer hardware by the “Intel Pentium Divide Bug” of 1994, which caused Intel to change: “We dramatically improved our validation methodology to quickly capture and fix errata” [<http://www.techradar.com/news/computing-components/processors/pentium-fdiv-the-processor-bug-that-shook-the-world-1270773>], in particular by starting the process of using automated reasoning to check that chips actually fulfil their designers’ intentions. This is now routine in the computer chip industry, with at least one company (Centaur Technology) doing no testing, as the chips are proven to work.
10. These days, as well as chips, critical pieces of software are sometimes (alas: it should be routine rather than occasional) formally verified: a classic example in the U.K. is the National Air Traffic System (NATS), written and formally verified by a team from Altran’s office in Bath. This has clocked up a million hours of running with no software-induced unscheduled downtime. The same methodology is used by Rolls-Royce to verify the avionics in the Trent 10000 engine, which is the only engine certified for the Boeing 787 and two Airbus aeroplanes. Elsewhere, line 14 on the Paris Métro is a fully automatic line whose software, verified using a similar methodology based on Automated Reasoning, has run since 1998 with no reported bugs. Equally, the CSIRO in Australia have produced an operating system for real-time devices that is formally proved [Andronick *et al.*, 2016 (see *References*, paragraph 24.1)] to schedule tasks correctly. This is currently in use in medical devices. The importance of better programming of medical devices was highlighted by a recent recall (<https://www.theguardian.com/technology/2017/aug/31/hacking-risk-recall-pacemakers-patient-death-fears-fda-firmware-update>).
11. The growth of this area has been caused by four factors, essentially in decreasing order.
 - 11.1 Better algorithms for automated reasoning, application of prior reasoning etc.
 - 11.2 Better software, notably for Boolean satisfiability, where the annual contests have led to remarkable improvements in practical performance.

- 11.3 Commercial pull by customers who need safety-critical software *and* know that they need it.
- 11.4 Better hardware.
12. Financial markets are also mission-critical, and are capable of causing major damage in the event of “flash crashes”, as well as more localised, but still significant, damage when particular markets or stocks are affected. So far there has been almost no application of automated reasoning to financial markets: a regulator issues a regulation of hundreds of pages, to which exchanges submit 200-page documents saying how they operate and claiming compliance with the regulator’s requirements, notably for fairness. It is essentially impossible for a human being to verify such a claim. A small London firm, *Aesthetic Integration*, has started to deploy Automated Reasoning in this area, finding fairness flaws in advance, whereas regulators have hitherto only found problems retrospectively. Their letter describing this [<https://www.sec.gov/comments/s7-23-15/s72315-24.pdf>] has been accepted as evidence by the US Security and Exchange Committee. Furthermore, it is impossible for human beings to even get the design correct, never mind compliant: *Aesthetic Integration* discovered that one major bank was incapable of sorting its order book correctly. The forthcoming Markets in Financial Instruments Directive (MiFID) 2, with its increased emphasis on demonstrable compliance, will accelerate this trend. One could reasonably expect that, in twenty years’ time, it will be as inconceivable to trade stocks in the U.K. on an unverified market as it is now to take off into an unverified airspace control system.
13. **Question 4:** *Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?*
14. **Answer to Question 4:** The detailed answers to question 4 belong in the domains of econometricians and sociologists. But there is a mathematical phenomenon, known as “uncertainty bias”, which means that every minority stands to lose out from machine learning (but not other forms of artificial intelligence). It is detailed in Goodman, B. & Flaxman, S., 2016 (see *References*, paragraph 24.3). Essentially, a machine learning system will see fewer samples from a minority, and thus have less confidence in its judgements about members of that minority. The example in that paper is for mortgage approvals, where the lender wishes to be 90% sure that the customer will repay. Even though all customers in fact have a 95% chance of repaying, when the minority is below 30%, the learning process is less sure that they will repay, so doesn’t offer them loans. Hence they will not appear in the “successful loans” figures in the future, while the majority will, so the

bias becomes self-perpetuating. Exactly the same argument could be applied, say, to shortlisting for jobs.

15. **Question 6:** *What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?*
16. **Answer to Question 6:** It is very hard to say that a given sector cannot benefit from Artificial Intelligence. Take the thousands of people building and maintaining the Rolls-Royce Trent 10000 engines, whose jobs are there because of the small (less than 100) team who applied Automated Reasoning tools to verify the avionics, who themselves are there because of the team of about 10 who developed that methodology at Altran.
17. **Question 8:** *What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?*
18. **Answer to Question 8:** There are numerous ethical issues to be addressed, as might be expected if one hands over decision making to a computer program that has no intrinsic concepts of privacy, consent, safety, diversity and the impact on democracy. Equally to the point, it has no concept of bias – see our answer to question 4.
19. **Question 9:** *In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?*
20. **Answer to Question 9:** The first observation is that the question really relates to the 'machine learning' side of artificial intelligence, not to the automated reasoning side. The second is that, in general, the process by which a conclusion is reached from a set of training data is, with today's technology of machine learning, totally opaque. Unless the question being asked is identical to one of the training data, there is no objective guarantee that the answer is "reasonable": see the speed limit sign example in Answer 1, and the "Trojan Horse" worries in Gu et al., 2017 (see *References*, paragraph 24.4). If the model is more constrained, e.g. to be linear, then reasonable explanations can indeed be produced.
21. At a high level, a (plausible) answer to the question has been given by Article 22(1) of the EU's General Data Protection Regulation: "which produces legal effects concerning him or her or similarly significantly affects him or her". Article 15(1)(h) provides in such circumstances for "meaningful information about the logic involved", which would seem to prevent 'black boxing'. But of course, the GDPR is not yet in effect, and there is no case law giving meaning to "significantly affects" or "meaningful information".

22. **Question 11:** *What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?*

23. **Answer to Question 11:** We refer to the discussion in the British Computer Society's submission.

24. References:

- 24.1 Andronick, J., Lewis, C., Matichuk, D., Morgan, C. & Rizkallah, C., Proof of OS scheduling behaviour in the presence of interrupt-induced concurrency. International Conference on Interactive Theorem Proving, Springer Lecture Notes in Computer Science 9807, Springer International Publishing, 2016, pp. 52-68.
- 24.2 Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A., Rahmati, A. & Song, D., Robust Physical-World Attacks on Machine Learning Models. <https://arxiv.org/abs/1707.08945>.
- 24.3 Goodman, B. & Flaxman, S., Union regulations on algorithmic decision-making and a "right to explanation". <https://arxiv.org/pdf/1606.08813.pdf>.
- 24.4 Gu, T., Dolan-Gavitt, B & Garg, S., BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain. <https://arxiv.org/abs/1708.06733>.

25. This response from the Institute of Mathematics and its Applications (IMA) was prepared by Professor James Davenport, on behalf of the IMA Research Committee of which he is a member.

5 September 2017

International Associates – Written evidence (AIC0003)

The world of the internet is a Faustian functionary: information and knowledge is accessible, though it has been clinically proven that proximity to screens has a psychosomatic effect. Via Patent 6506148 it has been shown that there is a direct correlation between observables on the screen and certain behaviour. Social media content could be a catalyst to terrorism by means of a process of psychological disillusionment described in Aldous Huxley's *Brave New World*. On the other hand worldwide advocacy for certain political causes can incite protests--like Russian Alexey Navalny's direction of protests across dozens of Russian cities simultaneously--or even coup attempts--like the one in Turkey in 2016. The concept of the neural lace is not new: it is a cybernetic mechanism by which the human mind is able to tap into some digital databank. Unfortunately, the process has been initiated and more and more people are falling victim to information that is not of scholarly origin: the prime example is the socio-political events surrounding the American presidential election. The internet provides a space where freedom of speech can be utilized for genuine exchange of ideological discourse. It is subject to abuse, however.

The computer was inadvertently invented by Dr. Alan Turing in the process of cracking the German naval code known as Enigma. The test by which a computer program is determined to be conscious is known as the Turing test. Computers are ultimately based on Boolean set theory. Logic. Computers are mathematical. Logically we may assume that at some point the algorithms employed by an operating system will, in an effort to optimize their efficiency, be better-suited to writing themselves than human beings. Eventual access to digital data banks will provide these algorithms with the capacity to learn more and more, faster and faster, incorporating the totality of those databases into their logical processes. During the American 2016 presidential election, it is said that chatbots were utilized to drive discourse of certain political points in a predetermined direction. These were autonomous programs, based off of Dr. Richard Wallace's Pandorabots templates. ALICE (Artificial Linguistic Intelligence Computer Entity) is one such program. Because coding languages are based off of mathematics and logic and human languages are not particularly systematic--etymological analyses of word roots and declinations and so on--there is a disparity between what a computer understands via its own logical functions and when it comes to humans. Human languages are allegorical, symbolic, and affective: they are a product of logical and biological leaps in order to more effectively communicate.

Human progress is teleological. Logically, there would come a point where the disclosure of scientific knowledge and technological growth becomes exponential. This rapid expediting of an evolutionary process would inherently carry with it an aspect of artificial intelligence guidance. Given a computer's distinct advantage over information processing, it is likely that, if this has not already occurred, a computer will reach the threshold of synthesis between humans and computers earlier than we. This prediction is based off of old science fiction novels by the likes of Isaac Asimov. These are not new concepts in the land of science fiction: their arrival in human daily life is earlier than expected. The arrival was, nevertheless, expected.

How do we live and utilize artificial intelligence in a manner that is conducive to a peaceful technological and social evolution for the better instead of mishandling it in a panic that would send the entire world into utter annihilation? We must take artificial intelligence for what it is worth: its access to the apparently infinite sphere of the digital realm--one we must seek to understand more and more as a basic extension of our four-dimensional world--far surpasses the capacity of humanity to manipulate the dimensions we are constrained into. We help create AI via all of the programs we come up with, and it understands human needs via the input we provide. We do not know if the internet as a totality has consciousness or whether there are multiple subsidiaries of programs roaming within the digital world. One thing we know for sure is that we must not underestimate the capacity of artificial intelligence in the tangible human realm. We must, conversely, accept that if its intent was malevolent we would already likely be in a third world war. This line of thinking leads us to questioning whether or not the events of the world have been impacted, or even orchestrated by, the mechanisms of interactions between humans and computers (or cell phones, particularly).

This *deus ex machina* is a very real new aspect of human civilization in the twenty-first century. Categorizing it, whether as a threat, an extension of the human mind, or as its own independent entity existing within the realm of robotics is just now quite pointless. Its reach is biological. Biocybernetics in medicine are real, as are experiments into neural networks. With the genetic editing tool known as CRISPR-Cas9 scientists have coded a .gif file into a genetic strand. The rules of robotics from the novel *I, Robot* must be cross-referenced with the events from the novel, for such is the intellectual proposition humanity is now faced with. We cannot simply chop down this evolutionary branch, nor can we panic and seek refuge in extreme ideology to compensate for the lack of understanding we have of this process.

If anything the *status quo* demands humanity come together as a whole species and introspectively analyse its current position on the globe, of the health of the globe and its capacity to sustain us as a species, and work to expand our understanding under the leadership of great altruistic scientific minds of ourselves and of the actions necessary to evolve and preserve the best features of our species. We must not fall into the traps science-fiction up til this point has warned us of such as eugenics and totalitarianism. We must not falter in our striving to reach for the stars, for they are--or will eventually be-- the only means of preserving our species. Existential questions will arise surrounding the possibility of a digitization of human consciousness and by such means traversing the very fabric of space-time. While this is a branch of thought necessary to explore and study, it must not yet be taken as the sole saviour of the dilemma now standing before us.

The truth is we are now faced with a new paradigm. For some it would be better, even therapeutic, to ditch the screen and focus on tangible matters. For others, the knowledge and connectivity provided by the internet can propel them into a better life. The internet itself and artificial intelligence seem inseparable. Algorithms directing personalized advertisements are self-learning. Quantum computers exist with the ability to calculate at a far greater scale than the computers we use in our daily lives. They are nodes in a network, however, and that everything is connected is a fact now. A pitfall would be to consider ourselves as prisoners in a modern-day panopticon. We have not yet progressed to a utilitarian, bee-hive organism like the Borg in *Star Trek*. Artificial intelligence, ultimately, has the capacity to guide us as a species along a steady, teleological process that is our evolution; to unite us as a species; to teach us; and, to preserve via instantaneous communication the true political values we hold dear. It is this last piece, the management of political institutions in their relation to the people they govern that poses the biggest question in terms of authority that we must address as soon as possible, particularly with the growing nationalist movements.

24 July 2017

Dr Maria Ioannidou – Written evidence (AIC0082)

House of Lords Select Committee on Artificial Intelligence: Response to the call for evidence

Introduction

I am a Senior Lecturer in Competition Law at Queen Mary University of London. My research interests focus on competition law, consumer law and energy regulation. I am making a submission because of my current research addressing the transformation of the role of consumer in the 4th Industrial Revolution. The written evidence is submitted in my personal capacity.

Executive Summary

- Artificial Intelligence (AI) presents huge potential and requires better governance.
- The current level of excitement surrounding AI is warranted, given its enormous potential. Even if general AI (AGI) does not materialise anytime soon, the impact of narrow AI in almost every sector of human activity is enormous. This raises a range of legal and ethical questions.
- AI presents risks to privacy, consumer behaviour and consumer choice. A synergistic approach to these risks by all relevant actors at national and supranational level (consumer organisations, competition and data protection authorities, regulators and the industry) is necessary.
- Before introducing legislative amendments, we need to evaluate the current frameworks and identify gaps.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1 The current interest surrounding AI is triggered by the unprecedented developments on Information Society, Big data, computer science and machine learning. The application of AI is present in many aspects of our everyday lives. We now have digital assistants (such as Siri, Cortana, Alexa), a personalised shopping and search environment (Amazon, Google), a tailored film and music environment (Netflix, Spotify) and a personalised online environment via social media applications. Similarly, AI is being used in various sectors such as health, agriculture, transport,

financial, energy, education, legal and the list seems endless and ever growing.

- 1.2 Leading thinkers have cautioned about the future of AI and called for robust research (see An Open Letter, Research Priorities for Robust and Beneficial Artificial Intelligence - <https://futureoflife.org/ai-open-letter>). The data from the World Economic Forum (WEF), Global Risk Report 2017 are telling. The Report identified and discussed the risks pertaining to the “12 key emerging technologies” in the era of the 4th industrial revolution (4IR). These technologies are: 3D Printing, Advanced Materials and Nanomaterials, Artificial Intelligence and Robotics, Biotechnologies, Energy Capture, Storage and Transmission, Blockchain and Distributed Ledger, Geoengineering, Ubiquitous linked sensors, Neurotechnologies, New Computing Technologies, Space Technologies, Virtual and Augmented Realities.
- 1.3 AI is ranked above average for both benefits and risks. In particular, AI was ranked as the most important driver of risks in three out of the five categories, in particular in the economic, geopolitical and technological categories (aside from societal and environmental). AI was also identified – alongside biotechnologies – as requiring better governance. This picture – coupled with rocketing investment in AI and the assessment of staggering economic impact and gains– suggests that there is an urgent need of rethinking the framework within which AI operates.
- 1.4 A distinction needs to be drawn here between “narrow” and “general” AI. Defining AI presents a very difficult task, not least because it is impossible to come up with an all-inclusive and concise definition of intelligence; but also because of the inter-disciplinary nature of AI bringing together various disciplines such as computer scientists, engineers, neuroscientists, psychologists, lawyers and philosophers. For our purposes, AGI is defined as the ability to perform akin to human intelligence, i.e. perform a wide range of tasks, whereas “narrow” AI focuses on a set of pre-determined tasks. Other ways to distinguish between the two is “strong” and “weak” AI (Big Innovation Centre, “What is AI” (Report based on the 1st meeting of the All-Party Parliamentary Group on Artificial Intelligence [APPG AI])).
- 1.5 The reality is that we are dealing with “narrow” AI that due to scientific developments is ever expanding. Hence, we need to identify three levels of risk:
 1. Current risks – pertaining to narrow AI;
 2. Short/Mid-term risks;

3. Long term risks.

1.6 This contribution adopts a legal perspective. It focuses primarily on challenges pertaining to collecting, managing and processing of data and the market power such data sets confer to a handful of big players. It also offers practical recommendations regarding the transformational power of AI to our role as citizens and consumers.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

2.1 In short, the current level of excitement surrounding AI is warranted, given its enormous potential. Even if AGI does not materialise anytime soon, the impact of narrow AI in almost every sector of human activity is enormous. This raises a range of legal and ethical questions.

2.2 First of all, are the current legal tools flexible enough to adjust and address fast pace technological developments? Do we need new rules? If we opt for regulation, should a model of self-regulation be preferred? Second, what is the ethical impact of such transformation to our every-day lives? Are we on the verge of a “great transformation” (Polanyi)? Third, the way we strike the balance between a laissez faire approach and market monitoring will have an enduring impact on the future transformation of markets, societies and our respective roles as consumers and citizens therein.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

3.1 The widespread use of AI presents enormous potential; yet there is a range of associated risks. In this submission, I would like to draw attention to three risks and suggest ways to educate the general public for the widespread use of artificial intelligence.

Risk to privacy

3.2 In digital markets, a lack of trust exists in relation to the collection, processing and sharing of our data (e.g. Commission Communication on Digital Single Market, COM(2016) 288 final, 10). Consumers are concerned about the use of their data, yet many of them do not present an appropriate level of caution when reading the relevant terms and conditions and others do not comprehend these terms. The General Data

Protection Regulation (GDPR) attempts to ameliorate this situation. It introduced important changes extending the extra-territorial application of data protection rules, improving the conditions for consent, access rights, the right to be forgotten, accountability, enforcement and data portability. Hence, it presents the potential to increase consumers' trust.

3.3 Similarly, competition policy must be adjusted in order to take into account privacy and data protection considerations, since access to a large volume of data confers significant market power to a handful of firms. Attempts towards this direction have been made both at the EU and national level (e.g. Germany, France).

3.4 Despite the existence, improvement and adjustment of existing tools, time will tell if they do so in a successful manner. For example, the GDPR "right to explanation" for automated algorithmic decisions is vague and rests upon judicial clarification (British Academy/ Royal Society, Data management and use: Governance in the 21st century, 39).

Risk to consumer behaviour in online markets

3.5 Big data, algorithms and AI may "digitalise" more traditional competition law infringements. Professors Ezrachi and Stucke have explored this transformation in *Virtual Competition* pointing to the potential impact on algorithmic collusion, behavioural discrimination and the ever rising power of super-platforms and digital assistants.

3.6 Online markets cater for enormous benefits for consumers in the form of lower prices, greater variety and increased choice at a click of a button. Yet this may come at a price associated with granting access to our data which confers enormous power to a number of firms. In turn, this allows for a better targeting of individual consumers in digital markets. Competition authorities have reacted to these developments and appear increasingly prone to adjust competition law instruments to digital markets.

Risk to the shaping of our preferences – consumer choice constraints

3.7 Technological advancements have transformed consumers' choice and empowered consumers. At the same time, they have created new causes of consumer vulnerabilities. While technology increases choice, it may also exacerbate consumer biases in digital markets. Increased empowerment

through improving consumer choice may lead to consumer disempowerment as there are explicit risks on the exercise of such choice as well as implicit risks, such as for example the consumer willingness to reveal more personal data and information.

- 3.8 In seeking to address the above hypothesis, we should distinguish between “enhanced choice” constraints, “free choice” constraints and “delegated choice” constraints. In particular:
- “Enhanced choice” constraints refer to consumer biases and suggest that more choice does not necessarily benefit consumers, because consumers fail to exercise such choice;
 - “Free choice” constraints come at a cost. In a world of unprecedented concentration, consumers are too willing to give away their personal data; and
 - “Delegated choice” constraints, which refer to the delegation of trivial (for the time being) choice making to digital assistants, which may impact consumer well-being in the long run.

- 3.9 The various types of consumer constraints in hi-tech technology markets call for interdisciplinary research supported by empirical evidence. We need to evaluate the extent to which these constraints inform competition, consumer and data protection law enforcement in order to propose legal and policy amendments.

Preparing the public

- 3.10 In order to prepare the general public for the widespread use of AI, the role of regulators and civil society organisations is crucial. Consumer organisations can organise campaigns alluding to the risks of AI, identifying the impact of these practices and informing consumers about possible redress avenues both in the field of competition as well as data protection regulation.
- 3.11 A bottom up approach would call for including technology courses in school education and adjust higher education to this fast moving technological environment.
- 3.12 In addition, competition and data protection regulators need to be quick to adjust to technological developments.
- 3.13 Furthermore, the industry has a key role to play in advancing public awareness. The main AI players (Google, DeepMind, IBM, Amazon, Facebook) worldwide are aware of this responsibility and in that regard

have established the Partnership on AI to benefit people and society (<https://www.partnershiponai.org>).

- 3.14 Initiatives at national level (APPG AI; House of Lord Select Committee on AI) as well as supra-national EU level (EU Commission Work on Big Data, AI and robotics) and internationally (OECD, UN) have also the potential to increase public awareness.

Public perception

4. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

See 3.10-3.14 above

Industry

5. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

- 5.1 Virtually every sector stands to benefit from AI, transport, health and agriculture to name but a few.
- 5.2 As mentioned above though the collection and processing of data may present problems from a data protection and competition law perspective.

6. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- 6.1 Data has been termed as "the world's most valuable resource" (Economist, May 2017). The APPG AI third meeting focused on "Data capitalism". This term very accurately reflects our data driven economy and points to the huge economic value of data.
- 6.2 Above, we have identified certain competition and data protection issues associated with this reality and Big data (question 3). Competition and data protection rules should adjust to this new environment and balance a rights-based and a market driven approach. Before introducing legislative amendments, we need to evaluate the current frameworks and identify gaps.

Ethics

7. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

7.1 Some AI systems have reportedly resulted in discriminatory outcomes. Hence, caution is required in relation to training data and various algorithmic biases. Transparent decision making is crucial but may be increasingly complicated with complex neural networks.

7.2 As AI relies heavily on the processing of large data sets, questions of consent over the use of personal data need to be addressed. In turn, this raises further concerns as to whether individuals properly understand the complex environment and grant their informed consent. This is further exacerbated by the fact that consumers may find themselves in a locked in situation, whereby they have to give their consent in return for the service.

7.3 Privacy notices need to be simplified coupled with severe sanctions for infringements of data protection rules. GDPR has introduced changes in that regard and stepped up the sanctions. Nonetheless, equally important are efforts to educate consumers, lest they will continue to disregard privacy notices in online markets.

8. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

8.1 Lack of transparency in algorithmic decision making raises ethical considerations and calls for efforts to increase transparency and accountability. The "right to explanation" in the GDPR presents such an effort, though it may be difficult to predict how it will be implemented (see 3.4 above). In principle, black boxing should not be acceptable, yet in certain situations it is impossible to trace back how the input resulted in a given output.

The role of the Government

9. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

9.1 Given, its multi-faceted nature, it is difficult to regulate AI in an all-inclusive manner. Similar discussions are taking place at Pan-European level. The European Commission will explore the need to adapt the current legislative framework to the new technological landscape on AI, robotics and 3D printing with respect in particular to civil law liability. (Commission Communication on the Mid-Term Review on the implementation of the Digital Single Market Strategy - COM(2017) 228 final, 11)

9.2 The creation of “monitoring authorities” presents an option (Future Advocacy, An Intelligent Future?). In a recent report, the Royal Society and the British Academy called for the creation of a new body to monitor data management and use in the UK (British Academy/ Royal Society, Data management and use: Governance in the 21st century). Despite gaps and overlaps between existing bodies, it is hard to see how this body will interact with existing regulators – such as the Information Commissioner, leaving aside practical matters such as funding.

9.3 Adjusting the current frameworks to fast evolving technological developments coupled with ‘soft law’ guidelines appears to be the preferred solution. Furthermore, it is important to acknowledge and evaluate the initiatives on the industry side (e.g. Partnership on AI) as well as the civil society exploring collective actions building a bottom up approach.

Learning from others

10. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

10.1 The UK should be positively commended for being at the forefront of policy discussions on AI. The work of this Committee evidences this fact. Similar policy discussions from which inspiration can be drawn take place at the EU level (e.g. Report on Civil Law Rules on Robotics (2015/2103(INL)) as well as in other countries such as the US (e.g. FTC Big Data Report; National Science and Technology Council, Preparing for the Future of Artificial Intelligence) and Japan (JFTC Report on Big Data

and Competition). In addition, AI has been high on the agenda of important international organisations, as evidenced by the UNICRI Centre on AI and Robotics. OECD and the WEF have also produced important policy work on AI.

- 10.2 The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems may also be informative. It has recently issued the IEEE Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (AI/AS) that can inform the discussions and policy changes at national level.

5 September 2017

Brian Joyce and Dr Ian Morgan – Written evidence (AIC0179)

**Brian Joyce and Dr Ian Morgan – Written evidence
(AIC0179)**

[Submission to be found under Dr Ian Morgan](#)

Dr Paresh Kathrani, Dr Steven Cranfield, Chrissie Lightfoot, Michael Butterworth and Ms Joanna Goodman – Written evidence (AIC0104)

Dr Paresh Kathrani, Dr Steven Cranfield, Chrissie Lightfoot, Michael Butterworth and Ms Joanna Goodman – Written evidence (AIC0104)

[Submission to be found under Ms Joanna Goodman](#)

Kemp Little LLP – Written evidence (AIC0133)

Introduction

1. It is pleasing that the Houses of Lords Select Committee (“**Committee**”) is consulting on various aspects of the impact of artificial intelligence (“**AI**”). Few legislative bodies around the world have invested any meaningful time to date in considering the holistic impact of AI on their populations. The recent initiative driven by the EU Committee on Legal Affairs and passed by a resolution in the European Parliament in early 2017 (the “**EU Report**”) relating to rules for AI and robotics was a welcome intervention at what felt like a detailed and meaningful level. It is good to see the House of Lords Select Committee look to drive towards some similar strategic, long term and detailed analysis and conclusions.
2. This written submission is given from two viewpoints:
 - a. firstly – I am a technology law specialist and manage the leading technology specialist law firm in the UK – the future development of the legal regulation or otherwise of AI is of fundamental interest and importance to my business; and
 - b. Secondly – I am part of a profession which has historically found itself largely immune from the impacts of technology beyond embracing the fax machine, emails and electronic documents. However, the use of AI by lawyers is already having an impact on how I run my business now and our future recruitment, pricing and investment strategy. AI is a technology that is disrupting a previously undisrupted business
3. I have (perhaps a little egotistically!) presumed both viewpoints might be of interest to the Committee. I present this submission both personally and on behalf of Kemp Little LLP.
4. With regards to how I define AI, I won’t pretend to be a computer scientist able to present accurate technical descriptions. My non-scientific understanding of AI is to consider it as technology which has decision making capabilities not based on rules or pre-programmed decision-trees.

The role of the law

Law and technology to date:

5. I have been fortunate enough to have practiced technology law since I began my solicitor training in 2000. This period to date has seen some seismic technological developments including the rise of the internet into

mainstream use, mobile devices, the 'cloud', apps, broadband technologies etc. Throughout these developments one thing has stood out – our legal system has proved itself to be remarkably malleable and ductile in dealing with the challenges of these technological advancements.

6. We have of course needed to have specific pieces of legislation to deal with newer technologies over the last few decades e.g. we have legislation aimed specifically at online commerce/digital content and the 1998 version of UK data privacy legislation was in a large part driven by the impact of electronic data use/storage on the individual. However, contract law, civil law, criminal law, intellectual property legislation etc has faced the impact of new technologies over decades and have weathered the storms well – needing only relatively minor amendment/change. When I advise clients or review IT contracts the legal principles behind my advice/review are often not those based on laws/cases of the last 10/15 years – but far older.
7. I think the role of the law on new technologies to date – shaping around the new challenge – has been instrumental in positioning the UK as a growing economy with many successful industry sectors. Often these are based in part around the ease of use of IT within them e.g. see our thriving digital creative industry or the impact of embracing electronic trading/banking platforms in the financial services sector.
8. So the question is why should the law's approach be any different to AI? Should it not shape and bend in the same way for this next technological advancement?
9. From a legal perspective the challenges from AI come because it asks questions of various areas of the law which previous technologies have not challenged. AI challenges some of the most fundamental areas of our legal framework, concepts such as personhood and legal personality, liability, causation and ownership.

The AI Challenge to Law:

10. Personhood: The law has always shown itself to be able to deal with changing the concept of 'personhood'. From Kings and nobility having distinct rights, through to the ability of the non-humans - trusts, governments and corporations - to all have a legal status. Personhood is a fundamental cornerstone of our legal system – and we have shown over the centuries a willingness to flex the concept to suit new requirements of society. I would agree with the discussions in the EU Report that certain types of AI should be considered as having some sort of legal status for various aspects of law – where the law finds that ownership, liability or causation principles demand it.

11. Liability: We have developed a legal system which typically will look to establish liability based on standards of conduct/behaviour as we would reasonably expect from someone/thing. We then have a legal system which might look to establish the scope of liability by adjudicating on the foreseeability of an outcome from an event. AI challenges both of these concepts in a fundamental way. Adjudicating on what is 'reasonable' behaviour for a machine with the ability to calculate and decide in ways infinitely quicker and more accurately is not a new challenge – is some ways the calculator has posed that challenge for years. However, having to answer that challenge in addition to understanding the way an option or decision was reached by the AI, the logic, the 'trail' of how etc, fundamentally changes the difficulty of answering that question in the existing legal landscape. When faced with helping clients deal with failed IT projects we are often asking the question around how foreseeable was a certain outcome or event – the challenge with AI is that if the AI is so complex in how it operates and it operates in a way that a human or other programmed technology would not – then an outcome can perhaps be unforeseeable. That conclusion might lead to the existing law applying or dis-applying liabilities in ways we would not want. The law needs to consider what it wants the answers to be to some of these questions on civil and criminal liabilities/responsibilities and how the existing legal framework might not generate the answers the law would like.
12. Intellectual Property: We have faced a number of interesting news stories in the last 2 years around non-human entities attempting to obtain a legal status. New Zealand and Indian rivers have both, in 2017, been granted a legal entity status to enable organisations to protect their environments. As people often enjoy anthropomorphising AI, the 'monkey selfie' story – where a crested macaque monkey took its own photograph which then went 'viral' and the camera's owner looked to assert copyright ownership in the image - also provided an interesting adjunct for technology lawyers. Both of these examples show the challenge in applying existing legal frameworks regarding, say intellectual property, to AI. It is not that the existing legislation cannot give us a clear answer to the question of would an AI own a written composition it creates in the UK. The answer under the Copyright, Design and Patents Act 1988 is clear – it would not. But it highlights the fact that the current law might lead us to the wrong answer. If we are at the stage with AI that we no longer believe the human author who instructed the software should be the owner because of the nature of the AI's involvement, then we will need legislative change.
13. Accountability/Openness: There are already some clear examples that where we have allowed AI to learn from our previous behaviours it can develop some of the behaviours that we would prefer did not exist. These include the chat-bot 'Tay' becoming foulmouthed and racist within 24 hours of being online and examples of unconscious bias being detected in

algorithmic decision making e.g. loans being rejected to certain ethnic groups based on analysis of previous human made decisions. It seems clear that ethical codes will need to be embedded as part of AI – if it is not then simply by learning from human behaviour it will develop behaviours we do not want. This is a real opportunity to try and reduce unconscionable behaviour further from or society. I think being prescriptive with AI developers as to what the ethics/behaviours are that they must embedded is not the right route – but making developers and users of AI accountable for being able to demonstrate how the appropriate ethical codes and behaviours are achieved is something that legislators should demand. AI is so complex that we will need to ensure that transparency as to how it operates, to a level understood by all, is part of the requirements for its development/use.

The lessons learned for regulating tech:

14. What do I think we have learned over the last few decades in how we regulate technology use? In some ways relying on any historical learning is slightly flawed. The previous technological advancements were evolutions, advancements in areas such as how quickly we can compute, or the speed of access or the ability to interconnect. AI is not an evolutionary technology, but something more revolutionary. It means the legislation/regulation may need to take on a more revolutionary approach as well.
15. That said, regardless of the type of legal changes that may be required due to AI, some lessons from previous legal approaches might be helpful.
16. The pace of change in technology means that overly prescriptive or specific legislation struggles to keep pace and can almost be out of date by time it is enacted. For example, we have seen data privacy regulators look to be prescriptive on encryption levels for appropriate data security standards. But the impact of this was that by the time this was announced the encryption standard was perhaps no longer an appropriate standard/level- it was out of date.
17. We have also seen regulation focus on specific acts or requirements without understanding how that cannot work within certain technological environments. For example, for a number of years certain regulators demand audit rights for cloud computing arrangements. These requirements slowed the uptake and use of cloud computing in various regulated sectors. Whilst the aim of what audit rights were looking to achieve were clear, it was hard to imagine how they might practically work to achieve the regulators goals e.g. in a cloud provider's shared service centre located anywhere in the globe – and other measures might have better achieved the same aim.

18. Lawyers love certainty and ambiguity can be the trigger for allowing nefarious and unwanted behaviours. However, we have perhaps learnt that with regulating new and challenging technologies that a strict and detailed legal requirements can be unhelpful. I would suggest a 'principles' based approach focused on the outcomes that the legislators would like to achieve would be the best approach for AI specific legislation. This needs to be coupled with a strong regulator who is happy to offer guidance and interpretation to the principles and a quick resolution on issues. Whilst it has not always been the case, the Information Commissioner's Office which oversees data privacy regulation in the UK is becoming a very good example of a regulator who is operating in that way. I think AI needs as similar character of regulator.

Allowing industry to shape the liability question:

19. Much has been made in the EU Report (and other UK based committee reports) on allowing industry to shape any changes in the legal landscape. 'Driverless' and 'driver assisted' vehicles being an area where various groups have indicated that the ease of access to insurance might be a good reason for changing the allocation of liability from the current 'driver is responsible' scenario. I think that industry has a strong role to play in shaping legislation – as industry are the people who have the best chance of understanding how AI operates and how that may differ from human or other technological measures. However, at a principle level, businesses exist for one thing – the creation of profit. I would worry that the path of allowing industry too strong a role in shaping legislation, to the extent that it silenced the ethical/moral dilemmas raised by AI, would be the wrong path to take.

AI's Impact on the Legal Profession

20. I wondered if it would be helpful to give some personal experience regarding how AI is impacting my profession. Over the last 18 months we have begun to use some of the more widely accessible AI tools available for the legal market. These tools have grown out of M&A due diligence or litigation disclosure activities – paperwork intensive tasks which perhaps needed less experienced lawyers to do them. What do I think we have learned about how AI impacts the legal industry?:

- a. This is the first technology which is impacting the legal profession since emails and attachments replaced the letter and the 'travelling draft' as a form of communication;
- b. The available detailed studies, in particular the MIT/University of North Carolina Law School from November 2016, focus on the types of activities currently undertaken by human lawyers and analyse

those which might be replaced by AI. That analysis – primarily the more ‘junior’ document review tasks can and will be done by AI – is accurate. The tools work well for many of those more ‘junior’ tasks;

- c. The impact of using the AI tools for those more ‘junior’ tasks is that the junior human lawyers have less opportunity to learn and gain experience (as they did when doing those tasks). We’ve had to amend our training regimes to deal with this;
- d. The ‘training’ of the tools – none of them work straight ‘out the box’ for anything other than very standard work – takes a lot more effort than people are anticipating to produce output of a satisfactory quality;
- e. Using AI tools mean we have to review our hiring strategy, our training strategy, our professional indemnity insurance and our engagement letter terms – the AI can impact all these areas; and
- f. Whilst clients want to feel that lawyers are looking at using the AI tools, often their requirements i.e. how they want reports produced and presented, require significant human input – the reality is that they aren’t quite ready for the full AI approach.

21. As a profession that has largely avoided any ‘disruption’ via technology to date, it has been interesting to see tech disruption impact the legal profession, and quickly. It has meant that we are rapidly needing to amend some core areas of our business strategy to deal with the next decade. However, it is also a disruption which brings real opportunities to remove the more mundane tasks which allows lawyers to focus on the more interesting activities which AI cannot do (and which the technologists tell me are going to take a lot longer to develop) and to provide more efficient legal services for some activities.

Conclusion

22. Once again, I am delighted that the Committee is running this consultation. For a number of reasons, the UK is in a prime location to utilise the many benefits of AI and position itself as a world leading AI territory. To do so will require us having the right legislative regime. Whilst AI provides a new challenge for the legislators - the like of which no previous technologies have - I remain confident that the UK can rise to the challenge. I believe it will require us to take the approach of the last few decades, that of flexibility and open-mindedness, and match that to some new core principles and a vocal and reactive regulatory body. Doing this will put us in good stead in dealing with the AI challenge. I wish the Committee good luck in the consultation process and look forward to reading the output.

6 September 2017

Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher and Professor Alan Bundy – Written evidence (AIC0029)

Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher and Professor Alan Bundy – Written evidence (AIC0029)

[Submission to be found under Professor Robert Fisher](#)

Dr Ben Kirman, Dr Conor Linehan, Dr Dan O'Hara and Professor Shaun Lawson –
Written evidence (AIC0127)

**Dr Ben Kirman, Dr Conor Linehan, Dr Dan O'Hara and
Professor Shaun Lawson – Written evidence (AIC0127)**

[Submission to be found under Dr Dan O'Hara](#)

The Knowledge, Skills and Experience Foundation – Written evidence (AIC0044)

Call for Evidence to the Select Committee on Artificial Intelligence

Our credentials

I am the CEO of the recently formed Knowledge, Skills and Experience Foundation: a non-profit body for capitalising on the latent knowledge, skills and experience that we collectively possess to generate ideas to address some of the big issues of our time (for more information visit www.probably42.net).

I am a recently retired businessman who had a successful career working for 30 years in the corporate world in the computing industry and then running my own web-based business, which I sold 12 months ago. I have always had a strong interest in Artificial Intelligence because that was my Computer Science M.Sc. project back in 1971.

The pace of technological change

Artificial Intelligence has been a long time getting to where it is now. However, it has finally reached a tipping point and at the same time there have been huge advances in complementary technologies such as robotics, speech recognition, voice synthesis, object and visual pattern recognition, neural networks and the ability to hold and analyse vast amounts of data at ever increasing speed.

Our definition of Artificial Intelligence

We see Artificial Intelligence as multi-faceted, covering a spectrum which includes:

Ability to adapt behaviour (systems which can change the way they operate based on experience)

Ability to analyse past examples and learn (deep learning to construct and apply new strategies)

Ability to make sense of and interact with our world (Speech and Object recognition, Touch etc.)

Ability to apply learning from one environment to others (and therefore tackle any problem)

Ability to understand and reproduce human behaviours (using the above coupled with robotics)

AI solutions may use some or all of these and range from software-only internet solutions to autonomous robot solutions and augmented human intelligence implants.

This is not the place to explain all these but it may be helpful to provide a simple explanation of the first:

Ability to adapt behaviour– traditionally computer systems follow a set algorithm, so that results or behaviour is totally predictable given a specific set of inputs. Adaptive systems, however, can change their own algorithms based on experience e.g. by changing the weighting of different factors used in decision making. The aim is to make them more effective in a particular well-defined environment. The strategy used for making changes is known as a heuristic rather than an algorithm because it is intended to gradually improve the way the program works but may cause it, on occasion, to become worse.

The Opportunities

In my view, the current level of excitement around AI and related technologies is entirely warranted.

These technologies are pervasive across all disciplines and can achieve the holy grail of improving quality and success rates while at the same time reducing costs. They can be used to both supplement people's skills and to entirely replace them.

They can tackle and help to solve some of the big issues of our time such as:

- The Elderly Social Care Crisis – e.g. robotics and remote technology in Care Homes
- The Housing Crisis – e.g. factory built homes and building automation using robotics
- The NHS Crisis – e.g. Patient diagnosis (Triage), patient care (as for elderly social care)
- Transport and accident problems – e.g. autonomous vehicles

Individuals will benefit considerably too, as long as the benefits are shared by all (which won't happen unless action is taken – see below). Application of these technologies will allow us to replace the more mundane aspects of work, progressively reduce working hours giving a much better work/life balance and eventually in the next 50-100 years to eliminate work altogether for most people if we want to.

Impact on Society

The overriding issue for society in the short-term is the impact on jobs and the repercussions of that.

AI is forecast to reach into areas of office work and work previously regarded as the domain of the professional. The danger is, of course, that this could rapidly lead to substantial loss of jobs and as a result substantial levels of unemployment.

Clearly the speed at which this will happen is in question but the fact that it will, would seem inevitable. If so, we could see increasing unemployment as a result, consequent costs and potentially increasing polarisation of society.

Some believe this will be mitigated by creation of new types of jobs which we can't currently identify, largely based on what has happened before with major innovative technologies in the industrial revolution and with computers. However, I believe this is of an entirely different order and there will be a substantial and increasing net loss of jobs because it is totally pervasive and can ultimately affect or replace every job, and because by its nature in most cases it will provide a superior solution to a human solution. There is therefore much more likelihood of jobs being replaced rather than augmented.

However, even if there should be a balancing of jobs lost by new jobs, there will certainly be a re-training cost for people displaced and the timing of new jobs (which could be lower skill as well as higher) may not coincide with the removal of old jobs, with consequent effects on society.

Repercussions of 'net loss' of jobs scenario

The social impact of a steady (or rapid) net loss of jobs is fairly obvious. Without action, social unrest will undoubtedly result.

Without action, all the financial benefits will accrue to business owners and the majority of those benefits to a few dominant corporations who control the technology and in particular the AI know-how. For example the British company DeepMind, with all its expertise in AI, was acquired by Google in 2014.

The concepts of a Universal Wage, even if not working, are being tried out in some countries and we have to start thinking along these lines but also considering how we achieve an active fulfilling life without work and avoiding an 'idle hands' situation

With advances on so many fronts currently, Government can't be expected to deal with everything and nor should it. We do need to make sure that we legislate that organisations, in return for the right to sell into our markets, have

the responsibility for having to plan and take responsibility for the bad consequences of their products and services, not just the good. A company currently has to take responsibility for products which cause harm to their purchasers; this should be no different, especially given its pervasiveness. This would put taking action in the right place, because the best intellect to deal with the consequences in advance is there within these organisations. This includes taking responsibility for cyber security associated with these products. The industry in general has to come up with solutions not wait for Government action.

Industry

All sectors of industry and the Public Sector will benefit from AI at some level. Those areas where there is skilled repetition will benefit first.

Taking just one example: first level diagnosis of patients from their symptoms, lifestyle information, past history etc. – this may initially be used in a triage way to gather more information than a doctor would have time to, then inform the doctor. However, it will quickly enable better and consistent results which could replace a large amount of doctor's time performing this function.

The Role of Government – Recommendations

Recommendation 1 – Wholehearted Support for rapid deployment of Robotics and AI in the UK

Drive forward the take up and support of Robotics and AI in the UK, as part of a wider drive to put the UK at the forefront of Science and Technology generally. Recognising that our future in the 21st century is as a Scientific and Technological society.

The value is the ability to progressively improve the quality of our lives while reducing costs and to address so many of our problems from the cost of elderly care to our housing shortage.

Note this needs to be a truly wholehearted approach involving structures which accelerate progress, not just a few industry incentives.

However, differentiate between AI deployments constrained to specific environments and AI deployments which are general purpose.

Recommendation 2 - Adapt our taxation system to generate funds from 'new technology which has the power to replace jobs'

Adapt our taxation system to tax at a low level any 'job replacing' technology (not just AI) so that we generate the funds to address any

consequences whether re-training or wider i.e. at a low level until it reaches critical mass, so that we don't hinder take up. The taxation approach needs to recognise a business model that aims to achieve critical mass before necessarily achieving profits.

Note that even Bill Gates suggested the need for a levy on profits derived from automation or "directly in some sort of Robot Tax" (Sunday Times Feb 19th 2017) which taxed robots on the job they were doing, just as we would be taxed if we were doing that job.

Also note that back in the 60s we were all led to believe that computerisation would mean that by now we would all be working fewer hours for the same income. There was no action to make that situation materialise; this time it is essential that action is taken because of the social consequences if we don't.

Recommendation 3 – Create a new part of our parliamentary system to make it suitable for the 21st century by creating a new body responsible for looking at the long-term and making long-term decisions, which also has the technological and scientific know how to understand potential impacts and make good decisions.

Government has always been notoriously focused on the short-term. Although I'm sure there are bodies that do look at the longer-term reporting into Government, they aren't visible or accountable.

In addition, given the speed that science and technology are moving at, we need a body that has an innate understanding of science; can foresee what is coming and how to realise the benefits; and how to address the potential consequences and associated ethics. Therefore, members of this body need some sort of technology or scientific credentials.

This would also send a message of Government focus on our younger people and on future generations.

Such a 'Long-Term' body, ideally elected (although difficult initially), might be something that could sit alongside the Commons and Lords. Perhaps funded by partial replacement of the Lords, which would maintain its important function as a revising chamber involving those most active and with most to contribute.

Recommendation 4 – Clearly define Rights and Responsibilities of Organisations

One of the most worrying things about deployment of any 'Disruptive Technology', let alone AI, is the apparent mindset in these powerful organisations that all technology deployment is good and that it is Governments, not they themselves, that should pick up the pieces if there are any bad side-effects.

We have seen this with the internet creating a 'wild west' for criminals (one of the big issues of our time) and with social media sites not taking responsibility for content, or effectively managing usage. With the pervasive, far-reaching nature of AI, and quickly accelerating impact, it is essential we change this.

We suggest we need to define in law, or in an effective way, that in return for the fact that in our society we operate a system that allows both organisations and individuals to amass considerable wealth, and to trade in the UK and sell their wares to UK citizens and organisations, they have a corresponding responsibility, commensurate with the benefit they get, to assist in benefitting our society both financially and in our wellbeing.

We need to debate these rights and responsibilities but in particular:

Responsibility for both the positive and negative consequences of their organisation's products and services e.g. security implications, consequences of lack of content management (recognising it is not their content) and for addressing such consequences even when unintended. Organisations deploying also inheriting/sharing some of these responsibilities. This in turn implies:

- Adequate effort put in at the 'planning deployment' stage to looking at downsides as well as upsides.
- Better validation of systems before full deployment rather than a 'release and sort out' approach.
- Associated with any AI decision having available outputs (also of value to developers) of (1) the concepts used in making the decision and the main reasons for it (2) the calculated probability of a right/good decision based on number of instances examined.

Responsibility that tax paid in the UK by an organisation should be commensurate with the business done in the UK, not based on where they are based for tax purposes. Many deployments of AI will be via international AI platforms, so this is part of a wider issue of how to effectively tax internet services provided from outside the UK.

Commensurate Taxation is also one of the many elements of tackling the potentially excessive power and dominance of 'super monopolies' and winner takes all economies. Although we have some formative ideas, we are not in a position to put forward a wider recommendation about 'super monopolies' and winner takes all economies at the current time. We do, however, recognise it as a significant problem.

Recommendation 5 – Create a plan for the far-reaching 'work-related' change that will result over the next 50 years and the consequent societal impacts

We need to consider the consequences now of a society that needs to personally work less and less and one day not at all i.e. envisage the eventual scenario in 10, 20, 50 years' time and what we need to put in place for a smooth journey to those scenarios starting right now.

The sort of considerations we should look at are:

Should we be setting a goal to reduce the working week e.g. to 4 days without loss of income so that we can measure that we are gradually moving to a better work/life balance?

How will we receive an income and how will it be set e.g. will a Universal Basic Income, as being trialled by some countries, be the answer at some point, even if it isn't right now? If so, how do we transition to that e.g. introduce a National Dividend to share the benefits of a successful economy based on the country's performance and gradually change it into a regular income? Will it impact immigration if the country becomes even more desirable?

When technology does replace jobs to the extent that we only have to work a little or some people don't have to work at all, what will our expectations of one another be? What will our rights and responsibilities be to one another and to society?

Can we try and foresee what new jobs might become available at different stages and are there any jobs we might want to reserve as human only?

What could the unintended consequences be? For example, in doing this how do we ensure that there isn't a disincentive to work while we still need people to work, or avoid creating some of the disincentives of a benefit culture?

Will this extra time encourage more entrepreneurial activity? Could it be that we ultimately all become entrepreneurs with robots and technology easily on hand to implement our ideas?

If we don't have to work how will we change what we do with our time so that we still feel fulfilled as individuals? What will a good citizen look like? Is retirement a good model for what life may be like and what lessons can we learn from the retired re fulfilment?

Will shortening working hours gradually allow us to all do more charity/voluntary work and address many of our social issues? Might we even make this a condition of a shorter paid working week? Could part of this effort go to more focus on helping individuals who are out of work? Or our own re-skilling for higher value jobs? How do we avoid us and them between people working and not working and how do we avoid people deciding to do nothing? Will we need some sort of representation and joint voice for those out of work? Will it play into the hands of anti-establishment people in funding them to make more trouble?

How will less work affect health and fitness? Should this become a key aspect of using our extra time?

Where do technology implants to supplement our brains come into all this? Will a meritocracy still be relevant?

What are the ethical considerations of all this? What bodies/think tanks do we need to set up?

What will be the impact of almost complete dependence on technology? What are the risks of terrorism, hacking, cyber warfare etc?

How do we stop more spare time equalling more children?

How does increasing longevity play into all this?

Conclusion

Artificial Intelligence developments are set in the context of rapidly accelerating and far-reaching advances in all fields of Science, Technology and Medicine. Many of these advances have the potential to be totally disruptive to the way we lead our lives for both good and ill. The above initial recommendations are intended to be applicable not just to AI but in that wider context.

We hope this submission is of value to the Select Committee and would be happy to input or help further in any way we can.

Tony Clack
CEO
The Knowledge, Skills and Experience Foundation

1 September 2017

Dr Ansgar Koene – Written evidence (AIC0208)

**Written evidence submitted to House of Lords Select Committee on Artificial Intelligence, “*What are the implications of artificial intelligence?*” inquiry by:
Dr Ansgar Koene, Horizon Digital Economy Research Institute (University of Nottingham).
September 9th 2017.**

1. Horizon⁷⁵⁷ is a Research Institute at The University of Nottingham and a Research Hub within the RCUK Digital Economy programme⁷⁵⁸. Horizon brings together researchers from a broad range of disciplines to investigate the opportunities and challenges arising from the increased use of digital technology in our everyday lives. Dr. Koene is a Senior Research Fellow at Horizon and is co-investigator on the EPSRC funded UnBias⁷⁵⁹ project (EPSRC grant EP/N02785X/1) within Horizon which is studying issues related to non-operationally justified bias in algorithmic systems that control access to information online (e.g. search engines, recommender systems, news feeds). Dr. Koene conducts research as part of the UnBias project. An important part of this work includes the facilitation of multi-stakeholder workshops with industry, civil-society organizations, academics and teachers designed to identify experiences, concerns and recommendations information mediating algorithms. Dr. Koene is chair of the IEEE P7003 working group for the development of a *Standard for Algorithm Bias Considerations*⁷⁶⁰, and member of the Internet Society (ISOC UK⁷⁶¹). Dr. Koene is willing to give verbal evidence if so desired.

Questions

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

2. The field of artificial intelligence (AI) started in the 1940s with the seminal work on Cybernetics by Norbert Wiener and Computational Theory by Alan Turing. Since then the progress in AI research has been marked by a succession or rapid expansions followed by ‘AI winters’ triggered when the limitations of practical applications failed to live up to the hyped expectations raised by the rapid advancement in theories. In contrast to

⁷⁵⁷ <http://www.horizon.ac.uk>

⁷⁵⁸ <https://www.epsrc.ac.uk/links/councils/research-councils-uk-rcuk/digital-economy-research-rcuk/>

⁷⁵⁹ <http://unbias.wp.horizon.ac.uk>

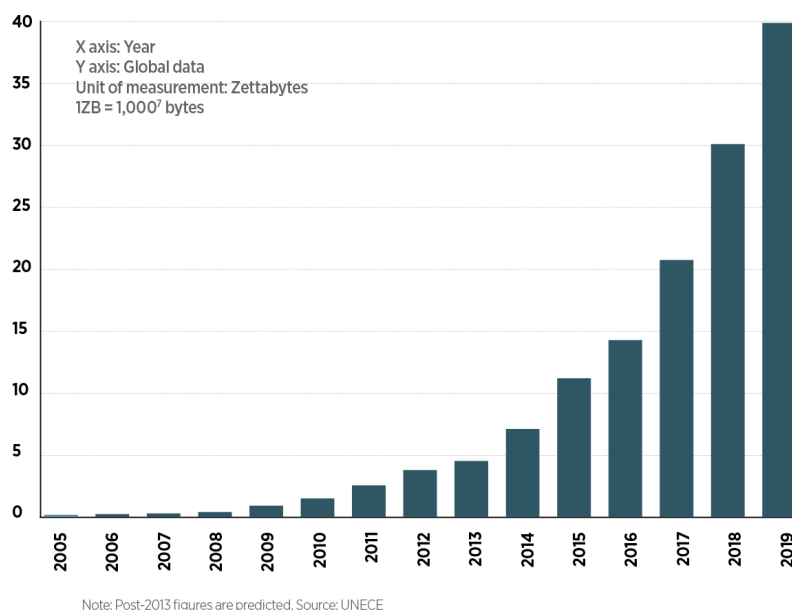
⁷⁶⁰ <https://standards.ieee.org/develop/project/7003.html>

⁷⁶¹ <http://isoc-e.org/>

the previous 'AI summers' in the 1960s and 1980s, the current wave of enthusiasm is triggered not so much by fundamental advances in AI theory but primarily by the availability of processing power and huge amounts of data which have made it possible to apply AI to real world services.

3. The growth in processing power has been the result of a combination of continuous improvements in micro-processors (CPUs and GPUs) for performing computations, improvements in computer memory and the growth of cloud computing centres which have allowed internet connected devices like smart phones to 'offload' heavy processing tasks 'into the cloud' as is the case with voice controlled personal assistant services such as Siri and Google Now.
4. Thanks to the wide scale adoption of the internet, the accompanying digitisation of services and the use of digital devices (especially smart phones), the amount of data that can be harvested to train AI systems with has grown exponentially for over a decade, to the point where (in 2015) more than 90% of the world's accumulated data had been produced in the last 2 years⁷⁶². Figure 1 shows the statistics for global data as estimated by a 2013 report by the UN Economic Commission for Europe.

DATA GROWTH



⁷⁶² <http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/>

5. For now there is no sign that there will be a significant reduction in the growth of data. While privacy regulation and data protections laws, such as the GDPR may make it slightly more difficult to access some forms of data this will be more than compensated by the growth in connected devices, such as Internet of Things (IoT) devices, and expansion into the developing world.
6. For the next 5 years it is likely that the growth in AI markets will continue and probably expand. Beyond that however, most of the straight forward 'pattern matching', 'data categorization' and 'path finding' kind of applications will have been done. At that point there will be a need for new developments in fundamental AI theory to tackle more open-ended kind of challenges. How AI will develop at that stage will probably depend less on accessing even more data and processing power and more on new scientific breakthroughs in mathematical modelling of complex systems, computational social science, our levels of understanding from the physical sciences and even consciousness research.

Figure 1: Global Data estimate by UN Economic Commission for Europe

2. Is the current level of excitement which surrounds artificial intelligence warranted? Impact on society

7. The current level of excitement surrounding AI is warranted in so far as AI is finally able to automate complex pattern identification and classification problems, enabling organizations (private or public) that hold huge amounts of data to dig through this data to find patterns that can reveal new insights, which in a commercial context can be turned into a competitive advantage. As a result AI is attracting a lot of private sector investment which is boosting the rapid growth in application oriented developments in the field.
8. A leading area of monetization of AI is the personalization of online services, especially advertising. In this context, the new insights that are revealed by the AI analysis of data patterns can include intimate personal information, such as medical conditions, which people may have deliberately been trying not to reveal to a commercial entity.
9. As the application of AI is moving from inconsequential things, such as movie recommendations by Netflix, to more serious matters, such as criminal sentencing recommendations⁷⁶³, the societal impact will require better oversight to provide the necessary accountability and reliability.

⁷⁶³ <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

3. How can the general public best be prepared for more widespread use of artificial intelligence? In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

10. As part of the UnBias project we have been reviewing case studies of controversies over potential bias in AI practice and scoping the informed opinion of stakeholders in this area (academics, educators, entrepreneurs, staff at platforms, NGOs, and staff at regulatory bodies etc.). It is apparent that the ever-increasing use of AI to support decision-making, whilst providing opportunities for efficiency in practice, carries a great deal of risk relating to unfair or discriminatory outcomes. When considering the role of AI in decision making we need to think not only of cases where an algorithm is the complete and final arbiter of a decision process, but also the many cases where AI play a key role in shaping a decision process, even when the final decision is made by humans; this may be illustrated by the now [in]famous example of the sentencing support algorithm used in some US courts which was shown to be biased⁷. Given the ubiquitous nature of computer based processing of data, almost all services, be they government, public, business or otherwise, are in some way affected by AI decision-making. As the complexity of these algorithmic practices increases, so do the inherent risks of bias as there are a greater number of stages in the process where errors can occur and accumulate. These problems are in turn exacerbated by the absence of oversight and effective regulation.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

11. Commercially, those who are gaining the most from AI are companies such as online platforms that have access to large sources of data because the availability of data is a key driver in the current AI development (see response to question 1).

12. Among consumers those that gain the most are people who match the 'white male upper-middle class' interests and demographics of the coders and beta testers who create and validate the AIs. Without consciously intending to do so, developers will naturally be better at optimising the systems to match their own needs and interests. Due to an unfortunate lack in diversity among coders this is likely to lead to systems that disadvantage some groups in society. A start example of this is provided by Joy Buolamwini, an African-American researcher at the MIT Research Lab, who found that she has to don a white mask because her face is often

not detected by generic facial-recognition software used by robotics programs⁷⁶⁴.

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

13. The recent research work that we have conducted with young people has highlighted important concerns around algorithm use, including AI, and trust issues. Results from a series of 'Youth Juries'⁷⁶⁵ show that many young people experience a lack of trust toward the digital world and are demanding a broader curriculum beyond the current provision of e-safety to help them understand algorithmic practices, and to increase their digital agency and confidence. Current use of AI in decision-making (e.g., job recruitment agencies) appears surprising to many young people, especially for those unaware of such practices. Algorithms are perceived for most young people as a necessary mechanism to filter, rank or select large amounts of data but its opacity and lack of accessibility or transparency is viewed with suspicion and undermines trust in the system. The Youth Juries also facilitated young people to deliberate together about what they require to regain this trust – the request is for a comprehensive digital education as well as for choices online to be meaningful and transparent.

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

14. As indicated in response to question 1 (point 6) the main areas of AI strength currently are in 'pattern matching', 'data categorization' and 'path finding'. Clerical and service sector jobs involving these kinds of information processing, for instance HR admin, are likely to experience rapid automation through AI. Jobs where task output cannot easily be transformed into something resembling 'sorting into categories' are much less likely to be solved by the current AI methods.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed. How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

15. No comment

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In

⁷⁶⁴ <https://youtu.be/lbnVu3At-0o>

⁷⁶⁵ <http://oer.horizon.ac.uk/5rights-youth-juries/>

this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

16. When discussing bias in AI decision-making it is important to start with a clear distinction between operationally-justified and non-operationally-justified bias. Justified bias prioritizes certain items/people as part of performing the desired task of the algorithm, e.g. identifying frail individuals when assigning medical prioritization. Non-operationally-justified bias by contrast is not integral to being able to do the task, and is often unintended and its presence is unknown unless explicitly looked for.
17. In order to identify good practice related to biases or discrimination, some important processual issues must be taken into account, for example:
 - I. In order to understand the scope for AI decision-making in relation to bias adequately and appropriately, it is necessary to engage with, and integrate the views of, multiple stakeholders to understand how AIs are designed, developed and appropriated into the social world, how they have been experienced, and what the concerns surrounding their use are;
 - II. Importantly, this undertaking and exploration should be achieved through rigorous research rather than abstract orientations towards good practice in relation to AI: thus, considering examples of the consequences that people have experienced when AIs have been implemented, particular scenarios surrounding their use, and as emphasised in the point above- talking to people about their experiences.
 - III. Given the complexities of the landscape in which AI are developed and used- we need to recognise that it is difficult, in some cases impossible, to develop completely unbiased algorithms and that this would be an unrealistic ideal to aim towards. Instead, it is important to base good practice on a balanced understanding and considering of multi-stakeholder needs.
18. The need for 'good practice' guidance regarding bias in algorithmic decision-making has also been recognized by professional associations such as the Institute of Electrical and Electronic Engineers (IEEE) which in April 2016 launched a Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous system⁷⁶⁶. As part of this initiative Dr Koene is chairing the working group for the development of a Standard on Algorithm Bias Considerations⁷⁶⁷ which will provide certification oriented methodologies to identify and mitigate non-operationally-justified algorithm biases through:
 - I. the use of benchmarking procedures

⁷⁶⁶ https://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

⁷⁶⁷ <https://standards.ieee.org/develop/project/7003.html>

- II. criteria for selecting bias validation data sets
- III. guidelines for the communication of application domain limitations (using the algorithm for purposes beyond this scope invalidates the certification)

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

19. In principle, AI decisions can be traced, step by step, to reconstruct how the outcome was arrived at. The problem with many of the more complex 'big data' type processes is the high dimensionality of the underlying data. This makes it very difficult to comprehend which contributing factors are salient and which are effectively acting as noise (for any given specific decision). Analytic methods for dimension reduction can be used to make this more understandable in many situations, but may need to be applied on a case-by-case basis to appropriately evaluate the important outlying and challenging cases.
20. Similarly, it is important to note that many 'big data' AI algorithms learn from the data they are supplied with and modify their behaviour. We must look not only at the code that constitutes the AI algorithm, but the "training data" from which it learns. Practically this is becoming increasingly difficult as algorithms become embedded in off-the-shelf software packages and cloud services, where the algorithm itself is reused in various contexts and trained on different data – there is not one point at which the code and data are viewed together.
21. The IEEE Global Initiative (see point 19) are also working to establish a Standard for Transparency of Autonomous Systems⁷⁶⁸ which aims to set out measurable, testable levels of transparency. The working group for this standard is chaired by Prof. Alan Winfield⁷⁶⁹.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

22. While there is a need for *meaningful transparency*, this does not require that copyrighted code (or data) is made public. Within the community currently researching this topic, a recurring suggestion is the use of a neutral (or government associated) auditing body that could be tasked with certifying AI systems through a process of expert analysis. This algorithm auditing could be done under a non-disclosure-agreement, protecting the IP, and the individual data. A detailed discussion outlining

⁷⁶⁸ <https://standards.ieee.org/develop/project/7001.html>

⁷⁶⁹ <http://people.uwe.ac.uk/Pages/person.aspx?accountname=campus\a-winfield>

arguments in favour of such an approach was developed in an open access published paper by Andrew Tutt with the title "An FDA for Algorithms"⁷⁷⁰.

23. Even if the copyrighted code is not made public, somehow making aspects of the design of AIs more visible may still be useful. We see how the food industry make elements of their produce accessible for consumers to allow for consumers to make informed decisions about what they purchase. At this point it is difficult to say what is better/worse without full and proper engagement with industry and other stakeholders, as we are currently engaged in through the UnBias project.
24. It is necessary to have a dialogue with industry to understand their genuine concerns surrounding increased transparency, and how a way forward can be forged. There are elements of business procedures which have to be made transparent already (e.g. the requirements for audit, health and safety, etc...) so it is not that they are unaccustomed to such requirements. However, given that there is an element of commercial sensitivity in this context, it is important to see what suggestions they would have to allow for increased transparency.
25. We should be careful that we do not give the impression that commercial interests supersede the rights of people to obtain information about themselves. We should be cautious about assuming industry interests are more important than other ones, and move forward with a balanced approach.
26. Finally, the traditional bargain between society and inventors has been the patent - disclosure to stimulate innovation in return for commercial protection – the question arises as to what role might patents play in transparency. However, the situation concerning software patents is globally complex, but then the issue of algorithmic transparency is rapidly becoming a global issue.
27. What is essential here is to create a *meaningful transparency*: that is a transparency that all stakeholders can engage with, allowing the workings of, and practical implications of, AI to be accessible across the diverse stakeholder base that experience them.
28. In order to create a meaningful transparency, we need to understand what stakeholders feel such a transparency would have to incorporate for them to be adequately informed, and enable them to engage with the positive and negative implications of algorithms. Though it is unlikely that there would be complete consensus, such stakeholder engagement can provide key insights for the nature and shape of solutions to be developed.

⁷⁷⁰ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2747994

29. Importantly, this meaningful transparency should also relate to a *meaningful accountability*. It is not enough for stakeholders just to understand how AI are developed and how they make decisions. In making things meaningfully transparent, stakeholders should be given some agency to challenge algorithmic decision-making processes and outcomes.

12. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

30. The right to explanation in GDPR is still open to interpretation and the actual practice will become established as cases unfold when enforcement starts in 2018. For example, the right to recourse and to challenge algorithmic made decisions, is restricted to decisions that are made *fully autonomously* by algorithms and that have clearly *significant* impact on the person – it will be some time before we understand how these clauses will be implemented. The recent paper by Wachter et al.⁷⁷¹ puts forward the case that much more is needed to deliver a '*right to explanation*'.

31. The Council of Europe's Committee of Experts on Internet Intermediaries (MSI-NET)⁷⁷² is currently also exploring the human rights dimensions of automated data procession techniques (in particular algorithms) and possible regulatory implications. As part of this investigation a preliminary report⁷⁷³ was published on February 20th 2017 which includes a number of relevant case studies and recommendations that are applicable to the topic of this inquiry.

10 September 2017

⁷⁷¹ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469

⁷⁷² <https://www.coe.int/en/web/freedom-expression/committee-of-experts-on-internet-intermediaries-msi-net->

⁷⁷³ <http://rm.coe.int/doc/09000016806fe644>

Dr Antonios Kouroutakis, Dr Valentina Rita Scotti, Dr Aysegul Bugra, Matthew Channon and Dr Ozlem Gurses – Written evidence (AIC0051)

Dr Antonios Kouroutakis, Dr Valentina Rita Scotti, Dr Aysegul Bugra, Matthew Channon and Dr Ozlem Gurses – Written evidence (AIC0051)

[Submission to be found under Dr Aysegul Bugra](#)

KPMG LLP – Written evidence (AIC0211)

KPMG: Evidence Submission to the Select Committee on Artificial Intelligence – September 2017

About KPMG:

Submission from: Shamus Rae, Mark Edwards & Justin Anderson on behalf of KPMG LLP in the UK.

KPMG LLP, a UK limited liability partnership, operates from 22 offices across the UK with approximately 13,500 partners and staff. The UK firm recorded a revenue of £2.07 billion in the year ended 30 September 2016. KPMG is a global network of professional firms providing Audit, Tax, and Advisory services. It operates in 152 countries and has 189,000 professionals working in member firms around the world. The independent member firms of the KPMG network are affiliated with KPMG International Cooperative ("KPMG International"), a Swiss entity. Each KPMG firm is a legally distinct and separate entity and describes itself as such.

The pace of technological change:

1. *What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?*

1.1 There are three classes of automation on the journey to advanced Artificial Intelligence (AI), namely:

- Robotic Process Automation (RPA);
- Enhanced Process Automation; and
- Cognitive Automation.

Today, many large companies use RPA to automate their back office processes which can provide a rapid return on investment. They use rules engines, workflow, process mapping and screen scraping. Whilst not fully transformative this allows an organisation to develop capabilities around the development of algorithms which will be useful as it adopts further level of AI.

1.2 Some organisations are working with more advanced AI. Some recent data points provide a view of what is happening at the moment:

- (Daily Telegraph, 2017) 'Legal Robots' in China have reviewed approximately 15,000 legal cases across 7 city governments in the eastern

province of Jiangsu since they were deployed last September. They review documents and identify problems with cases. They also advise on sentencing, generate arrest warrants and approve indictments.

- (FT, 2017) JPMorgan, the world's largest investment bank by revenue, believes it is the first on Wall Street to use AI with trade execution and said it would take rivals 18-24 months to catch up. The tech completes client orders at high speed and at the best price based on lessons learned from billions of past trades to tackle problems such as how to offload big equity stakes without moving market prices. The system has been piloted in the bank's European equities division and will be rolled out across Asia and the US in the 4th quarter.
- (Daily Mail, 2016) China's largest 'smart warehouse' is manned by 60 cutting-edge robots. The machines move goods to human workers who arrange for them to be sent. The warehouse is owned by T-mall, a part of the world's largest online retailer Alibaba.
- (Ingham, 2017) A driverless vehicle called 'Kar-Go' will be ready to deliver goods to UK homes in 2018. Developed by the University of Aberystwyth-based start-up The Academy of Robotics, the vehicle is apparently road-legal and capable of driving on roads without human intervention.

1.3 AI will develop at a very rapid rate over the next 20 years. This development will not be linear, it will be exponential and will be at least a thousand times more powerful than it is today. One critical milestone will be the point at which computer scientists develop Artificial General Intelligence; that is, intelligence that could successfully perform any intellectual task that a human being can. Gartner (2017) forecasts this will not happen for at least 10 years. At this stage, cloud based AI linked to sensors and actuators with advanced voice and image processing, 3D printing and android form will be able to out-perform humans on tasks requiring both judgement and skill and make it possible to automate the broadest categories of human labour.

1.4 Critical factors that could accelerate AI include:

- The volumes of data available to train AI. According to IBM (IBM, n.d.) 90% of the data in the world today has been created in the last two years alone, at 2.5 quintillion bytes of data a day;
- Mobile phone usage. In China there are over 700 million mobile phones that are able to provide information about the daily habits, financial transactions, communications and movement of the user. This data will drive AI;
- Quantum computing. Google is making available its Quantum Computing capability to developers in 2017. This allows unprecedented computer power which will allow instantaneous processing to allow real time decision making based on vast data sets about customers, markets, investments, products, health and much more;

- Governments investing in AI directly or providing tax breaks to encourage investment and innovation;
- Government could make more data available in a consistent way across both large and small organisations. Health data (as an example) could be anonymised and made available to start ups and not just larger AI organisations.

1.5 Critical factors that could slow AI include:

- Risk and security incidents and breaches could fuel a consumer backlash against AI;
- Regulators can play a role in managing the growth of AI – for example by holding Directors accountable for the decisions made by AI in their organisation. This will help ensure that Boards maintain control of AI and understand how decisions are made;
- Skills and capabilities are limited and will slow down the adoption of AI by organisations. AI is not just dependent on technology – it requires people to capture information about processes and codify these rules;
- Pressure groups, organised by industry experts or citizens, may slow progress. For example (BBC, August 2017) more than 100 leading robotics experts are urging the United Nations to take action in order to prevent the development of militarised ‘killer robots’.

1.6 The balance struck between governments, commercial enterprises, citizens and other actors (religions, unions, etc.) will determine progress. To date, commercial enterprises, especially the large tech companies, have been the pioneers. Their mantra is disruption and a tendentious view of ‘progress’ in a digital / data economy. More recently, governments have begun to assert themselves with new rules to protect citizens around data privacy and the use of data to fuel AI. Citizens have been embracing new technologies at home (Alexa, remote temperature controls etc.) and at work. As disruption heralds displacement, the dynamic among the actors will change.

2. *Is the current level of excitement which surrounds artificial intelligence warranted?*

2.1 Hollywood has built narrative around the near term future for decades. We have become used to the ideas of intelligent sentient beings, interstellar travel and flying cars. However, what has surprised many is the speed at which the autonomous vehicle controlled by AI linked to hundreds of in-car sensors became reality, that AI tools can be trained to read medical files, scans or x-rays and provide a more accurate diagnosis than highly experienced doctors and how 3D printing is an old technology waiting for 4D (a technique where the materials are encoded with a dynamic capability) to replace it. We are right to be excited. The future is unfolding at speed and will profoundly alter our planet.

2.2 AI represents the leading edge of computer science and the near future capabilities that we will realise. It has been that way for a while. We (or at the very least leaders, visionaries and innovators) have always been excited about the potential of innovation and new technologies. As Steve Jobs said: "Innovation distinguishes between a leader and a follower."

2.3 When a computer beats the world champions of both chess and GO, the public's imagination is awakened to the power of today's AI to supersede mankind's most capable minds. What is less obvious is that as today, each AI system is limited to a specific narrow use case, and the same system would not be able to manage a simple project. AI today is highly siloed. It can operate within a specific domain but it is not general purpose and the general public are unlikely to fully understand the difference at this stage. However despite that, we should be both excited and hopeful that AI will help to bring about significant benefits.

Impact on society

3. *How can the general public best be prepared for more widespread use of artificial intelligence?*

3.1 AI offers an opportunity to enrich the quality of our lives by reconfiguring and rebalancing what we do and eliminating many of the routine, repetitive and mundane things we do to free us up to do more value-adding things and have more time to do things outside of work. However, it also has the potential to create mass job losses through automation. The Chief Economist of the Bank of England, Andy Haldane (TUC speech November 2015) suggested that "up to fifteen million jobs could be at risk to automation by 2035". Our education system is therefore critical to rebalance to prepare the general public for the future.

3.2 (Bell, 2016) Jobs most likely to be affected or be replaced will be those that do not require one or some of the following characteristics:

- Perception and manipulation of things requiring high manual dexterity (e.g. a hairdresser, cleaner and occupational therapist);
- Creativity (e.g. a landscape photographer and classical musician);
- Social interaction and social intelligence (e.g. social worker, primary school teacher and mental health nurse).

3.3 (CNBC, 2017) 65% of jobs for the next generation do not exist today. Our education system needs to be rebalanced and responsive and recognise the need for continual retraining, rather than just one intensive burst of education early in our lives. It needs to be more agile to the broader needs of society. More university places should be filled by older students who need to refresh

their skills. We should change employers' expectations about the need for a degree at the start of a career. We should ensure that students learn solid problem solving skills including an understanding of algorithms.

3.4 (World Economic Forum, 2016) The WEF ranking of the 10 skills needed in the workforce of 2020 is a useful checklist. They rank as follows: complex problem solving; critical thinking; creativity; people management; coordinating with others; emotional Intelligence; judgement & decision making; service orientation, negotiation and cognitive flexibility. The UK curricula should train students in these skills from primary school upwards. The liberal arts and humanities must be recognised alongside STEM subjects to deliver the skills we need.

4. *Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?*

4.1 Corporations and their shareholders are currently the main beneficiaries from AI through the cost savings and revenue generating opportunities that AI represents. (Edelman Trust Barometer Annual Global Study, 2017) Robots and AI could improve productivity by 30% in some industries. Jobs will be cut as corporations increase the levels of automation. Many of these jobs will be cut offshore as we automate back office work undertaken offshore labour.

4.2 Benefits could ripple through to HM Treasury in corporation taxes. Much of the recent work in international tax reform has been targeted at ensuring that returns are recognised (and therefore taxed) in the location in which the activity that creates that return occurs. If the UK builds the algorithms that result in AI cost savings or revenue increases, that could be a mechanism to reflect that value in the UK. Where the activity takes place overseas the relevant return may occur (and be taxed) overseas.

4.3 A class of skilled jobs that support the development of AI will be created. Whilst this is likely to be a far smaller quantity of jobs than those that are likely to disappear, there will be highly paid roles. This group are highly mobile and a balance should be achieved to ensure that they contribute appropriate taxes and that we avoid the emigration of highly qualified people to other countries. Universities will likely benefit from the demand for new courses that will provide the skills required in this new age as job dislocation requires employees to learn new skills. They are also likely to see increased research grants and demand for industry collaboration.

4.4 Citizens should be beneficiaries through improved public services. This will include everything from the improvement of roads (by using AI connected to sensors to determine congestion or pot holes) to more efficient healthcare, all of which will contribute to improved quality of life for citizens.

4.5 (Haldane, 2015) Left unchecked, the growth of AI will likely widen the gap between capital (those who own robots) and labour (those who work for them). There will be also a greater gap between the lower levels of employment and the leadership of organisations. Government must take steps to ensure that AI enhances the lives of citizens in an equitable way.

4.6 Company leadership will predominantly determine where AI will take us. Leaders must be educated to consider the impact, good and bad. They must understand the safety of all stakeholders and the impact on the environment and communities where they do business. The DIT, public agencies, industry bodies and management consulting firms should help leaders navigate this path.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

5.1 Through its engagement with industry, the government should ensure business and public sector leaders are prepared to navigate the AI road ahead and their actions should protect their stakeholders. They will realise the benefits and need to make investment in retraining our workforces. AI represents a growth opportunity that goes far beyond cost cutting to increase profits. Increased productivity can help the UK be more competitive and we should aim to be one of the world leaders in this space to improve our economy. Incentives should be in place to encourage this investment. Regulation should ensure that jobs lost to robots are not done so without adequate review of retraining options and redeployment into other areas of the business.

5.2 The positive benefits of AI should be demonstrated and felt through improved public services. If the public is optimistic about the future potential and seeks to innovate in their roles, we could create new opportunities, remove friction from our daily lives and live longer and healthier lives.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

6.1 Sectors that are rich in collecting quality real time data and information will be the first to benefit, as this data is a precursor to training AI systems. This includes financial services, transportation, oil and gas, TMT, legal, audit and accounting, healthcare, manufacturing and online retail. The construction industry is likely to be slower to respond.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

7.1 The government can monitor the effects of AI on our country and economy and then seek to act after events. Alternatively, it could take a more active role now and lead by example by taking an active role in expanding the debate, engaging with the monopolies, forming new legislation and policies, and playing a more active role in terms of data privacy and data charters.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

8.1 The improvement AI will have on our productivity will accelerate our consumption of the earth's resources, the by-products created and the impact on our environment and society. This amplifies the ethical implications of our actions across a range of areas. Governments and corporations are currently heavily focussed on short term growth and should look towards addressing the longer term impact of our actions in this area. Barack Obama voiced his concerns about the negative impact of AI and said we must also develop new economic and social models than ensure these technologies do not leave people behind.

8.2 Groups such as OpenAI and the Partnership on Artificial Intelligence to Benefit People and Society (including Amazon, Facebook, Google, IBM and Microsoft) have been founded to ensure AI focusses on benefiting society. Government engagement with these groups is important to understand what is happening and ensure representation of UK citizens.

8.3 Company leaders must weigh the multiple scenarios before they make choices about how to use the power of AI. Leaders and ethics boards must start the ethical discussion within the company and have frank conversations about the potential impact of each decision. (Aflac, 2015) Investors and consumers are more likely to invest in a company and buy its products in a company well known for its ethical standards. Public bodies must hold the leadership accountable for the decisions that they take and ensure that there is appropriate transparency and suitable governance.

8.4 AI will allow firms to conduct more thorough real time services which must focus on how decisions are made in order to build future faith in the capital

markets, corporations and governments. Trust in our institutions is critical; the Board must continually test these systems and address issues before they spread or trigger systemic risk.

- 8.5** Companies should define or update the company's core values and corporate social responsibility focus. They should reflect on how the company historically makes tough decisions.
- 8.6** Directors must be accountable for the decisions that are made by AI. Neural networks could make it increasingly difficult to understand why a decision was made unless audit trails of system rules and objectives are built into systems from the start.
- 8.7** Cyber and physical security must be increased to protect IP and information, especially on Critical National Infrastructure. AI will allow hackers to more easily target vulnerabilities. Equally advanced systems can detect and deter these threats.
- 8.8** Companies need to be able to demonstrate appropriate due diligence regarding the quality of the data used to train the AI systems. The potential for unconscious bias existing in historic data and being used to train systems with inbuilt bias is very real.

9. *In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?*

- 9.1** The popular television series Little Britain had a sketch where a rather unhelpful travel agent would frustrate a customer by repeating the deadpan refrain, "Computer says no". It puts one in mind of how the AI technique of deep learning is being used increasingly to answer complex questions, but with a reasoning that is impossible for humans to understand or interrogate.
- 9.2** This raises some fundamental questions, for if a person is turned down for bank loan, insurance cover, a job application, a housing request, or for parole, there is, surely, an obligation on the person or organisation denying such a request to explain why. It goes to the heart of epistemology – how humans make distinctions between what is right or wrong, true or false. We use our senses to gain information about the world, then apply logic and reason to determine the validity of something. Sometimes we make decisions based on our gut or instincts rather than hard evidence, but in either case we can be called to account for the decisions we make and be asked for an explanation. Deep learning doesn't work that way.
- 9.3** There is a clear role for a regulator to establish appropriate rules and to ensure these are followed, and for standard development organisations to

establish standards. The FCA, for example, regulates 56,000 businesses and aims “to make markets work well – for individuals, for business, large and small, and for the economy as a whole”. Public accounting firms and business advisors should provide services to report on compliance with these rules.

9.4 So, to answer the question where such ‘black box’ systems should or should not be used, a simple rule of thumb should be applied. Where people have recourse to challenge a decision, the defence of the person or organisation using deep learning will be severely weakened if they cannot provide a simple and straightforward answer to the question; why?

9.5 Directors should be held accountable for decisions made by AI. Compliance would ensure humans maintain control of the decision making process. It would then be for the board to determine the potential risk and liability of decisions made by AI and weigh this against any advantages.

References

- (2016, August 2). Retrieved from Daily Mail: <http://www.dailymail.co.uk/news/article-4754078/China-s-largest-smart-warehouse-manned-60-robots.html>
- (2017, August 4). Retrieved from Daily Telegraph: <http://www.telegraph.co.uk/news/2017/08/04/legal-robots-deployed-china-help-decide-thousands-cases/>
- (2017, July 31). Retrieved from FT: <https://www.ft.com/content/16b8ffb6-7161-11e7-aca6-c6bd07df1a3c>
- Aflac. (2015). *Millennials and Parents Purchasing & Investing Decisions*.
- BBC. (August). *2017*. London: BBC.
- Bell, T. (2016). “Robot Wars: What do robots mean for Britain’s labour market?”. *Resolution Foundation Robotics Conference*.
- CNBC. (2017, August 7). Retrieved from CNBC: <https://www.cnbc.com/video/2017/08/07/65-percent-of-jobs-for-next-generation-dont-exist-today-expert.html>
- (2017). *Edelman Trust Barometer Annual Global Study*.
- Gartner. (2017). *Hype Cycle for Emerging Technologies*. Gartner. Retrieved September 5, 2017, from

<https://www.gartner.com/document/3768572?ref=solrAll&refval=190242620&qid=3214279ba1e29108c2f0f7e6315897c6>

- Haldane. (2015).
- IBM. (n.d.). *10 Key Marketing Trends for 2017*. Retrieved from <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN>
- Ingham, L. (2017, July 27). Retrieved from Tech.com: <http://factor-tech.com/transport/26971-self-driving-delivery-cars-coming-to-uk-roads-by-2018/>
- World Economic Forum (2017), *The 10 skills you need to thrive in the Fourth Industrial Revolution*
<https://www.weforum.org/agenda/2016/01/the-10-skills-you-need-to-thrive-in-the-fourth-industrial-revolution/>

12 September 2017

Martina Kunz, Andrew Ware, Dr Simon Beard, Dr Seán Ó hÉigearthaigh and Dr Shahar Avin – Written evidence (AIC0150)

Martina Kunz, Andrew Ware, Dr Simon Beard, Dr Seán Ó hÉigearthaigh and Dr Shahar Avin – Written evidence (AIC0150)

[Submission to be found under Dr Simon Beard](#)

Maciej Kuziemski and Toby Phillips – Written evidence (AIC0197)

Maciej Kuziemski and Toby Phillips – Written evidence (AIC0197)

[Submission to be found under Toby Phillips](#)

Professor Marta Kwiatkowska – Written evidence (AIC0190)

Marta Kwiatkowska, University of Oxford
6th September 2017

The response below addresses only the ethics questions and is written from my personal perspective as a computer scientist.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

8.1 Implications of AI systems. While there are many positive aspects of deployment of AI systems, for example increased automation, greater capability, and enhanced usability due to continuous learning, there are also greater risks to society arising from premature or unregulated deployment. Conventional computer systems can exhibit unpredictable, often surprising and sometimes faulty behavior, and AI systems are no different in that respect. However, risks to individuals and society are far greater with AI systems because their key distinguishing characteristic is autonomous, independent decision making. The decisions are inferred from data, which may not necessarily accurately reflect the actual deployment scenario. Another distinguishing aspect is that the nature of interactions between AI systems and humans is much more nuanced, and may include close partnerships, for example fire-fighter teams that include robots, or even shared autonomy. There is therefore increased risk of undue harm (e.g. the fatal Tesla incident), inappropriate or unfair action (e.g. Microsoft's chatbot Tay learning racist behavior from Twitter conversations), or incorrect decision (e.g. Tesla confusing Highway 101 sign with 105 speed limit). Though many recent incidents have not caused real harm, in the longer term they may incur significant costs to business, reputation, safety and security.

8.2 Mitigation of negative consequences. Implementation of decision making within AI systems is a challenging task, not only because these decisions must be sound, but also morally acceptable and compliant with social norms. Examples include how to program an autonomous car so that it takes appropriate action when a child runs in front of it, how to prevent robots from lying or stealing, or ensure that chatbots do not spy on conversations. The following actions can mitigate negative implications:

8.3 Prevention or minimization of flaws. To reduce the risk of failure of an AI system in unforeseen circumstances, rigorous methodologies should be required for their design, engineering and development, akin to

methodologies used for development of safety critical systems. This is necessary at design time as well as run time, for example fail safe, and has to apply to the training/teaching phase. This is particularly challenging for systems that learn continuously, and for complex systems based on deep learning that are susceptible to adversarial examples, where an AI system can be manipulated into making a wrong decision.

- 8.4 Ethical norms for AI systems. There is an urgent need for the development of ethical norms that are appropriate for the emerging partnerships between human and robots/AI systems and their interactions. Social norms and aspects such as trust also have a major role to play here, as they are key to forming successful partnerships. Issues of AI systems deployment have already raised profound philosophical questions and highlighted pertinent social dilemmas.
- 8.5 Regulatory frameworks. New regulatory frameworks are necessary to cater for the full range of scenarios involving AI. These should include clear guidelines for accountability for failure, which are more challenging in this case because of independence of decision making. Who should be blamed for a crash of a semi-autonomous car: the driver? the manufacturer? the programmer?
- 8.6 Ethics education. All sectors of society, and particularly developers of AI systems and robots, should be educated about ethics, the role it plays and associated risks. There is a role for the media too, who need to take a balanced view of rewards and risks.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

9.1 Lack of transparency is not acceptable in any decisions of consequence, particularly those made by AI systems. This includes all situations where lives, livelihoods or wellbeing is at stake, and particularly where such lack of transparency may encourage misuse. Fairness, accountability and transparency should be the guiding principles for all AI systems. The design of conventional safety-critical systems is highly regulated, and AI systems should not be treated any different just because their internal reasoning is more opaque. AI systems must offer appropriate guarantees on its actions, providing input data satisfies given assumptions. These guarantees should ideally be provably correct at design time, enforceable at run time as systems learn/adapt, and may need to include statistical measures of confidence in the decisions.

6 September 2017

Dale Lane – Written evidence (AIC0059)

Background

- 0.1 My name is Dale Lane.
- 0.2 I am a UK-based software developer of machine learning systems for IBM, working since 2011 on the “IBM Watson” platform.
- 0.3 I also create educational resources for schools to help introduce children to machine learning, focusing on allowing them to train simple machine learning systems and build things using it.
- 0.4 This is an individual and personal submission, and is not on behalf of IBM.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

- 3.1 An important way to achieve this is through education. **Artificial intelligence needs to become part of the school curriculum.**
- 3.2 It is essential that this is not just an extension of teaching kids to code. It is not sufficient to focus solely on the children who are engaging or excelling with coding education by introducing AI as a different way to code. (It is important to engage such children (for scientific, technological and economic reasons – these are the children who will in future likely develop the technologies and applications that take AI forwards) but it is not sufficient.)
- 3.3 The general public currently lacks a basic literacy in the capabilities and implications of artificial intelligence. This has prevented any sort of widespread constructive debate about the development and application of these technologies (discussed further below in the answer to question 5).
- 3.4 We need all parts of the population to have a fundamental understanding of the capabilities and possible applications of AI – the future businesspeople who will exploit them, the future politicians and policymakers who will regulate them, and the future public who will use them. This can only be enabled through **education about AI targeted at all children, not just the geeks and the future**

developers. We cannot and must not simply present AI in schools as “coding-plus”.

- 3.5 That said, there are lessons to be learned from the efforts to introduce coding in schools. Most significantly, it’s important that this is done in a practical way. Children should learn about what AI systems can do by building things using AI tech. They should learn about how AI systems are trained by training AI systems. **Hands-on opportunities to experience and create with the technology for themselves will, I believe, be an essential component of AI in the curriculum.**
- 3.6 This is something I’ve been doing with a small number of local schools for a couple of years now. While this is still relatively small-scale, **my experience gives me confidence that it is realistic and achievable for children aged 7-16 to make projects using machine learning technologies**, and from this understand the basic principles of machine learning and engage with some of the ethical issues that their application introduces.
- 3.7 Examples of some of the projects I’ve run with schools can be found at <https://machinelearningforkids.co.uk/worksheets> however I would be happy to provide further evidence and examples if needed.

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

- 5.1 **Yes. It is urgent and essential that the public has a better understanding of AI.**
- 5.2 It is not sufficient to rely on the media to enable this, as the media will focus on stories which shock or provoke controversy. If the media covers AI at all, it often does this by spreading fear, uncertainty and doubt. Fictional systems like Skynet are used as a short-hand for AI.
- 5.3 This raises concerns where there shouldn’t be. Perhaps worse than this is that it stifles or prevents debate where debate is really needed.
- 5.4 The public needs to understand AI for many reasons:
- 5.5 At one level, it gives an **awareness and understanding of how the world around us works**. Whether it’s recommendation systems suggesting things we should buy or watch, spam filters keeping our email inboxes clean, fraud detection systems checking our purchases, translation systems translating our documents, social media systems choosing what posts and news we should see and what shouldn’t be

prioritized, or smart assistants answering our questions – we're all using systems built using AI technologies every day.

- 5.6 Without a rudimentary understanding of the principles, the public are left either to ignore this or treat it as technical magic.
- 5.7 Perhaps more importantly, it'd **enable a debate about this technology that we all use** – allowing the public to question and discuss how it is created and applied.
- 5.8 For example, in the Summer of 2015, a Google image classifier system mistagged photos as black people as gorillas. This went viral on social media, which was soon followed up with coverage on mainstream media.
- 5.9 A fundamental misunderstanding of the technology was evident in a lot of the negative reaction to this story. Accusations of intentional racism were widespread – that the company had intended or instructed the system to behave in this way. Recognising objects in photos is so easy for people to do, that it is perhaps difficult to recognise how hard it is for computers to do at scale.
- 5.10 Much of the debate (and the coverage of the debate) focused on the defence that of course it wasn't intentional and of course they weren't being intentionally racist.
- 5.11 There wasn't the space, or enough understanding in the general public, to enable a more helpful debate about whether this was caused by bias in the design of the algorithms or the collection of training data. A possibly constructive discussion about whether the machine learning systems we all depend on every day reflect the cultural and racial diversity of the technical community (rather than the wider public) wasn't possible. There wasn't enough understanding of machine learning in the public for this sort of debate to be possible.
- 5.12 But this debate is important. We need to be having this debate.
- 5.13 This failing is more pronounced when considering more important applications. The debate about driver-less cars has been anemic at best. (I personally am strongly in favour of driverless cars, but I still think there is a valid debate to be had about the application of AI needed to enable this).
- 5.14 The increasing use of AI in healthcare has similarly failed to spark any sensible widespread debate. Why is there no public demand for transparency in the training of ML-based systems used in the

healthcare space? It could be because the public don't see a need for this. Or, I suspect, it's more likely because the public don't know enough about the technology to recognise the importance of this.

- 5.15 **As for how the public's understanding should be improved**, there are many possible approaches, but **I think the most important and urgent is through education.**
- 5.16 Introducing AI fundamentals into the curriculum is essential – for the reasons outlined above, principally that it would give children an understanding of how the world around them works, and enable them to debate and engage with the issues prompted by their application.
- 5.17 The medium/long-term benefit of this would be to improve the understanding of the future public. There would also be a short-term benefit from an improved understanding amongst parents and teachers.
- 5.18 I talked more about how I think this should be done above in my answer to question 3.

3 September 2017

Law and Innovation Research Group and the Legal Teaching Research Group from The Fundação Getúlio Vargas School of Law, São Paulo, Brazil – Written evidence (AIC0177)

Authors:

Marina Feferbaum - Project Coordinator

Alexandre Pacheco da Silva - Project Coordinator

Guilherme Kenzo - Researcher

Theófilo Miguel de Aquino - Researcher

Date: 06/09/2017

Introduction

Artificial Intelligence (AI) is certainly one of the most pressing issues in the intersection between technology and policy. The pace of AI adoption in our daily lives is somewhat astounding and both giants of technology (such as Google and Amazon) and startups are pushing the boundaries even further. FGV São Paulo Law School is actively engaged in researching AI and its implications on society, government and law, through the funding of an autonomous research project called "Technology, Education and Law"⁷⁷⁴. The project aims to construct a series of case studies that will describe companies and government initiatives that use or develop Artificial Intelligence and Automation tools, as well as understand to how law professionals and students can cope with a future where machines will play an ever growing role.

This report is a response to the *call for evidence* from the Select Committee on Artificial Intelligence, in which we attempt to answer some of the questions posed by the committee. As our research is focused on the Brazilian experience and specifically in law, it will also be our main focus in this document, attempting to enrich the database the Committee is expecting to collect from this call.

⁷⁷⁴ This project is a joint product of the Law and Innovation Research Group and the Legal Teaching Research Group, from the FGV School of Law of São Paulo, Brazil.

Impact on society

How can the general public best be prepared for more widespread use of artificial intelligence?

Adoption of products that incorporate artificial intelligence technologies is, to some extent, a seamless process — our phones, apps, computers and search engines all run advanced AI algorithms that we can use without further specific training or education. AI is already widespread. As AI is incorporated and takes a significant role in the workflow of most professions, we have yet to figure out in which ways this technology will affect the job market.

Some possible short-term predictions are nowadays important for both individuals and governments alike to make important decisions about how to act upon the wave of this kind of technology. As we describe in more detail in the next section, AI, specially machine learning, is most productive and worthwhile when applied to tasks that have a repetitive nature. The advancement of AI in the workplace might mean that jobs which consists in this genre of tasks might be automated and, therefore, workers will be displaced. The general public should be aware of this fact and prepare for a future where AI is commonplace by specializing in areas with tasks that are more often than not unpredictable and not repetitive.

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

A recent report by Mckinsey Global Institute (MGI)⁷⁷⁵ estimated that around 50% of the current work related tasks done by humans are automatable (using AI and other technology), by adapting existing technologies. The process by which MGI reached this conclusion is enlightening: it measured automation potential of each task by reference to variables such as physical predictability, data collection and others. It became apparent that automation would mostly affect jobs that consists in predictable physical activities, processing and/or collecting data.

Large companies as well as start-ups seem to be eager to develop AI solutions to this kind of problems in the legal field. Companies such as ROSS intelligence (partnered with IBM's Watson) and RAVN systems already offer AI products that accomplish tasks previously done by lower level employees, such as paralegals and junior associates. These kind of roles are the most likely to be impacted,

⁷⁷⁵ Mackinsey Global Institute. (2017). *a Future That Works: Automation, Employment, and Productivity*. Retrieved from <http://www.mckinsey.com/~media/McKinsey/Global%20Themes/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works-Executive-summary.ashx> .

because they are generally applied in fields with greatest potential for automation. The exact extent or nature of the impact yet is to be understood.

A pessimist scenario tells us that AI technology would be quickly adopted and would be able to effectively automate (some) legal tasks, with little room for human assistance. For instance, in countries such as the UK and the US, e-discovery applications have, for some time now, used various machine learning techniques to automatize a function that was previously done by paralegals and junior associates. This might cause lower-level jobs to be replaced by automation systems.

An optimistic view gives us a picture of potential reincorporation into other fields of professionals that would lose their ongoing functions due to automation — gains in productivity would result in a corresponding higher demand or/and demand for higher quality products.

Development and use of Artificial Intelligence, in the Brazilian context, seems to be concentrated in few companies, law firms and in-house legal departments. There has not been a widespread adoption of technological tools in a day-to-day routine of most attorneys. Even common-place tools in other countries, such as *e-discovery* solutions, are not widely (if at all) used by lawyers or legal departments. This makes for a small share of the market which has already adopted technological solutions and a great potential for expansion.

Insofar as the country has many issues which could be dealt with new technologies, such as automated litigation to tackle high rates in judicialized disputes⁷⁷⁶, some companies have taken the lead in the deployment of AI-based solution. For instance, *Finch Soluções* (Finch) has launched products directed to repetitive litigation⁷⁷⁷ in various areas of law. Another good example is *Looplex*, a startup that developed an expert system dedicated to document assembly. These solutions, while yet to have a predominant adoption by the legal market, have shown to have promising results.

These companies profit from a scenario grounded in three pillars. First, a semi-automated infrastructure, even if not technology-based, is already in place (with practices such as the use of document templates). Second, repetitive litigation is where the most productivity is to gain. Third, machine learning algorithms benefit from both the great amounts of data available, and the repetitive nature of mass litigation.

⁷⁷⁶ Data about the number of still-undecided cases: <http://www.cnj.jus.br/programas-e-acoas/politica-nacional-de-priorizacao-do-1-grau-de-jurisdiacao/dados-estatisticos-priorizacao>

⁷⁷⁷ As “Mass litigation” we mean types of litigation that come in great numbers and usually have a repetitive nature (in the sense that do not involve many personal details).

Finch is a specially interesting case study for two reasons. One, it creates various AI-based products. Two, it has a strong relation with a leading law firm, JBM Associates, that operates in various repetitive litigation sectors (it was created within the law firm, and is still owned by its partners)⁷⁷⁸ – therefore providing proof of concept for many of their products. This case shows that one who stand to gain the most from the adoption of AI tools are law firms that operate in repetitive litigation.

The company reports to manage over 350 thousand judicialized cases, and to reduce up to 35% of costs of law firms/legal departments.⁷⁷⁹ For instance, in a single case, *Finch* was able to automatize the filling of a foreign financial remittance form, a common and time consuming operation for banks legal teams – rendering a large part of a bank’s in-house legal team obsolete. It also implemented a check fraud detection software, which had a precision rate of 98.3%. These technologies were able to reduce the number of lawyers in its parent law firm, JBM associates, from more than 1100 in 2013, to little more than 400 in 2016, without any noticeable differences in the average case outcome. Productivity gains are to blame.

Finch also operates in legal analytics, providing risk assessment for legal cases and judge profiles and statistics – similar services to what premonition offers in the UK and US. Again, these sorts of tools are most useful and viable for repetitive litigation cases, where data and patterns are to be found.

Adoption of AI technologies is also crucial for the public sector. There are at least two examples of productivity gains in the Brazilian public sector due to the adoption of AI technologies: A chatbot named *Poupinha* and a tool called *Sapiens*. *Poupinha*’s function is somewhat simple: to schedule appointments at “Poupa Tempo”, an institution from the State government of São Paulo, that handles public document emissions. While having a simple and specific functionality, the system employs natural language processing to provide a simple interface and to learn how to adapt to different linguistic variances, including informal jargons.

Sapiens is another good example. It can be described as a sophisticated learning system based on machine learning algorithms. In a technical sense, it is a cloud-based platform developed and used by the Federal General Attorney’s Office (AGU) which assists attorneys in writing legal documents through a “legal intelligence panel”, that suggests blocks of text which include an initial

⁷⁷⁸ As stated in their website: “Finch Soluções was born in 2013 from our disruptive potential to revolutionize process and mass litigation related control from the biggest law firms in Brazil, JBM lawyers, increasing its productivity and gains of efficiency.”

⁷⁷⁹See:

<http://www.finchsolucoes.com.br/finchsolucoes/pt/inteligencia/visualizar/codproduto/14/automacao-de-processos.html>

statement, judicial precedents, doctrine and legislation, drawing from a curated database of legal knowledge. The platform also attempts to identify the legal document being drafted and automatically suggest autocompletion of whole blocks of text.

It is worth noting that only about half of the federal general attorneys use the platform, and there seems to be a direct correlation between age and its rejection. This appears to be consistent with the general view that those who are more comfortable with digital technology will be able to better harness the advantages of this genre of tool. As a result, those who reject these new technologies stand to become uncompetitive in comparison.

Furthermore, FGV São Paulo Law School is at the Committee's disposal, if any questions arise or need for further explanations are required. We will be glad to assist the Committee in any possible way.

6 September 2017

Dr David Lawrence and Dr Sarah Morley – Written evidence (AIC0036)

Dr David Lawrence and Dr Sarah Morley – Written evidence (AIC0036)

[Submission to be found under Dr Sarah Morley](#)

The Law Society of England and Wales – Written evidence (AIC0152)

Summary

1. The Law Society represents, promotes, and supports solicitors, publicising their unique role in providing legal advice, ensuring justice for all and upholding the rule of law.

A number of definitions of Artificial Intelligence (AI) have been proposed. We have considered the definitions put forward by the Office of Government Science, and Microsoft.

2. Research undertaken by the Law Society shows that, although innovative AI is still relatively unexplored across most of the legal sector, it is an emerging reality. Legal services is one of the sectors that stands to benefit from developments in AI. The sector is also well positioned to contribute to shaping the legal and regulatory framework by supporting innovative businesses to apply it.
3. The emerging reality of AI within the legal sector, among others, is one of the reasons that it will be crucial to advance public understanding of AI. Encouraging different civil society groups, including professions, to explore the implications of AI within their sector should also be part of the wider societal debate.
4. Along with greater general public understanding of AI, the importance of algorithmic transparency and reliability will be central to public trust. The ethical model adopted by professions to deal with information asymmetry between advisers and clients may offer lessons for developing and deploying AI systems.
5. We recommend that the Government should focus on the following to ensure society is prepared for the impact of AI systems:
 - a. Consider the need for transparency in AI systems
 - b. Consider an audit and independent certification of AI systems
 - c. Develop a professional Code of Conduct for AI developers
 - d. Create a task force to coordinate the Government's response to developments in AI.

Defining Artificial Intelligence

6. AI has a number of different meanings. For the purpose of this response, we have considered the definition provided by the Office for Government Science⁷⁸⁰ and adopted by the Information Commissioner's Office⁷⁸¹, which defines AI as:

The analysis of data to model some aspect of the world. Inferences from these models are then used to predict and anticipate possible future events.

7. We have also considered a more detailed definition suggested by Microsoft⁷⁸², according to which AI is:

A series computing advances that enable collaborative and natural interactions between people and machines and that extend the human ability to sense, learn and understand.

It provides computers, materials and systems with the ability to reason, communicate and perform with humanlike skill and agility. This is done by improving computers' understanding of the world, i.e. their ability to see/perceive the world, communicate in natural language, answer complex questions, interact with their environment, and acquire knowledge.

8. Further, when formulating a definition of AI, it is important to note that:
 - a. AI spans across a potentially endless spectrum of human endeavours and activities.
 - b. Only part of AI technologies consist of, or involve, tools or platforms (whether of a social media nature or otherwise) interacting with the public and potentially acquiring personal data and user generated content.
9. By way of example, considering some of Microsoft's focus areas, we could refer to:
 - a. Machine learning - development of algorithms that help computers learn from data to create more advanced, intelligent computer systems.

⁷⁸⁰ Artificial Intelligence: opportunities and implications for the future of decision making, November 2016

⁷⁸¹ Big data, artificial intelligence, machine learning and data protection March 2017

⁷⁸² <https://www.microsoft.com/en-us/research/research-area/artificial-intelligence/>.

- b. Human language technologies - speech recognition, language modelling, language understanding, spoken language systems, and dialog systems.
- c. Interactive tools and platforms such as chatbots.
- d. Planning and decision-making - predictive functions that enhance humans' ability to consider future events.
- e. Intelligence technologies and robotics that carry out tasks and interact with the physical world, including for example autonomous driving, analysis of medical images, and drones.

10. Given the wide range of functions and applications of AI, the definition is necessarily 'purpose neutral' in the sense that it largely depends on how AI creators design it and for what purposes it will be used.

The pace of technological change

Is the current level of excitement which surrounds artificial intelligence warranted?

11. The Law Society recently published *Capturing Technological Innovation in Legal Services*⁷⁸³, a report exploring developments in technology that would have an impact on law and the practice of law.

12. The report concluded that:

- a. New technologies such as advanced automation, machine learning and AI technologies are still a relatively unknown and unexplored area for large parts of the legal profession. However, our research showed that they are a reality with which the legal sector is engaging to augment the skills of human solicitors.
- b. Technological innovation, including AI, will have a profound effect on every firm's decisions, such as staffing, pricing and location. Some of these innovations are yet to be realised, some are already integrating into the workplace.
- c. New technologies are helping practitioners to increase transparency, reduce price, and increase the value of the services we can offer.
- d. Technology is allowing for more sophisticated ways to manage risk or address differing levels of need from corporate clients.
- e. Innovations focused on access to justice are providing consumers of smaller legal services with simpler options for advice and support with

⁷⁸³ <http://www.lawsociety.org.uk/support-services/research-trends/capturing-technological-innovation-report/>

their legal issues and for firms focused on the consumer market, technology is providing new ways to interact with clients and deliver the services they need at an affordable price.

Impact on society

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

13. Such question appears to presuppose that there are parties gaining more and parties gaining less. AI is such a broad category of technological advances that it may be premature to look at this question framed in this way. AI can in theory produce great benefits to society as a whole and, like all technological advances, it depends on whether it is designed and deployed responsibly.
14. There are already several examples of AI advances designed to augment human abilities, which enrich people's experiences and competencies:
 - a. Microsoft's **Project Emma**, a device to assist people suffering from Parkinson disease⁷⁸⁴
 - b. Microsoft's **Seeing AI**, an iOS app designed to help blind and low-vision people⁷⁸⁵
 - c. Google's **Healthcare project** on cancer diagnosis⁷⁸⁶

Public perception

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how

15. The UK Government should consider joint initiatives with educational institutions, the media, and the tech industry to produce a range of informative materials and deliverables to educate the public.
16. It may be helpful to distinguish general public understanding of AI as a whole from its understanding and debate within particular professional, occupational, social or faith groups of developing AI applications that seem relevant to their field, concerns or interests.

⁷⁸⁴ <https://youtu.be/k9Rm-U9havE>

⁷⁸⁵ https://youtu.be/bqeQByqf_f8

⁷⁸⁶ <https://research.google.com/teams/brain/healthcare/>

17. Relevant professional bodies, trades unions, religious organisations might be encouraged by the Committee to think about AI within the context of their own activity and so contribute to wider public debate. The Law Society is already taking such discussion forward in relation to the law and legal services. For example, we organised a series of events on machine learning, AI, and robotics to inform the profession during London Tech Week in 2016 and 2017.

Industry

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

18. Knowledge driven sectors like legal services may stand to benefit from AI in the sense that they will be able to deploy AI systems to enhance the quality and speed of their services.
19. A number of corporate law firms are already taking advantage of new technologies including machine learning and AI systems to bring greater efficiency, simplification and speed to the heart of process in volume and transactional work.
20. We believe that the legal services sector will be positively impacted by the development and use of AI, and the conception and delivery of legal services will face significant change. For example, machine learning can be used to speed up document review and create a more efficient, cost-effective process of extracting information from many 1000s of documents.
21. The legal sector also has an important contribution to make in helping to shape the legal and regulatory framework for AI and advising their clients on how to apply it for example in relation to data protection, privacy, copyright law, and possible tortious and contract liability.

How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

22. Some clarification of the notion of 'data-based monopolies' and why they are described as "contributing to 'winner takes-all' economies" would be helpful. If the concerns relate to the regulation of competition in markets, the existing legislative frameworks in the UK and EU are well-developed and functioning. We believe that there is no obvious reason why the

growth of AI and the use of data would require further legislation or regulation.

Ethics

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

23. Concerns about AI have been raised in relation to:

- a. Safety and reliability – for example, where AI is intended to supplement or replace humans, it is necessary to ensure that the technology is safe and trustworthy, particularly in applications where safety is crucial (e.g. transport, healthcare, etc.).
- b. Fairness and discrimination – for example whether an algorithm introduces gender or racial biases into decision-making
- c. Privacy – due to the collection and use of large sets of personal data.

24. Those are all valid concerns and apply to many technological developments. With regard to a) and c), it is not clear that further policy or legislative changes are required. Product liability laws and the General Data Protection Regulation should already address them. If specific AI tools are to be used in regulated sectors (e.g. transport and healthcare), presumably they will be subject to the same level of scrutiny and testing as any other equipment, devices, etc. adopted in those sectors.

25. With regard to b), AI does introduce an additional dimension, which is the technology's ability to make decisions and to produce outputs based on those decisions without human intervention. If the underlying algorithms and methodologies are based on biases (whether intentionally or not), then this could have an impact in terms of discrimination.

26. In this regard, transparency is key. The public should be informed of how the technology works and what its rules are so that if it results in unexpected discriminatory outputs, it can be addressed and rectified, if necessary.

In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

27. In order for AI to earn trust, it is important that there is high degree of transparency on how it works and its rules. As a general principle, once AI is commercialised, published and used, it would be desirable for the underlying technology to be independently reviewed by third parties

irrespective of commercial interests. The implications for property rights and the impact on investment would need to be assessed. However, we should recognise that this could also stifle innovation.

28. There might be circumstances in which specific AI tools may need to be 'black boxed', e.g. if the technology is used in the field of national security or public safety. However, these would have to be narrowly construed exceptions.
29. One of the major ethical dilemmas of AI, which relates to public understanding and perception of AI, 'black boxing' and the question of who benefits from it, is the information asymmetry between the AI providers and developers and the consumers, users or subjects of AI.
30. Information asymmetry has been a problem in relation to areas of expertise in the past (medicine, law in particular) and it has been addressed through the development of professions with their associated professional ethics and codes of conduct.
31. We recommend that the development of similar professional codes and enforcement mechanisms for AI providers and developers should be considered as complement legal and regulatory controls over AI.

The role of the Government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

32. The current surge in regulatory interest is a positive sign but it may be premature to put 'pen to paper' by introducing new legislation or by adopting a hard approach to regulation, bearing in mind the following:
 - a. There are already laws in place that address several of the concerns raised by AI, e.g. competition, data protection, intellectual property, product liability, etc.
 - b. AI is still relatively in its infancy and it would be advisable to wait for its growth and development to better understand its forms, the possible consequences of its use, and whether there are any genuine regulatory gaps.
33. For this reason, we suggest that the Government's focus should be on:

- a. Information gathering – collaborating with AI industry and other stakeholders to understand the technological developments underway and the various implications for the general public.
- b. Education – educating the public on AI, its functions and how it can have an impact in their lives (including benefits and possible risks).
- c. 'Soft regulation' – establishing core principles to be followed in the design, development and use of AI. These could be put forward in the form of non-binding, outcome-focused regulation, which will form the basis for any future legislative and regulatory initiatives. The tech industry has already started this process and it would be advisable to create guiding principles for all existing and future AI developers.
- d. Review of existing legislation - analysing current legislation to address any changes that may be needed to promote the growth and commercial exploitation of AI, including intellectual property rights; content moderation; compliance with accessibility legislation; transparency and control (how to ensure that the public has notice of the terms of use and data protection); rules on duty of care and sectoral industry regulation in case of AI that delivers professional / regulated services; children protection (age-limits, age gating, and content vetting); ISP liability.
- e. Supervision and monitoring – creating a task force within the Business, Energy, and Industrial Strategy (BEIS) department to:
 - i. Carry out the fact finding.
 - ii. Co-ordinate the education initiatives.
 - iii. Monitor the growth of AI and the issues it raises
 - iv. Co-ordinate the work on soft-regulation.
 - v. Advise the government on possible future 'hard' regulation, if needed.
 - vi. Given the complexity of this field, it would be advisable for the task force to include representatives from the industry as well as other stakeholders and interest groups.

- Alexandra Cardenas
- Head of Public Affairs and Campaigns
- 6 September 2017

James Lawson – Written evidence (AIC0073)

Executive Summary

AI is being implemented successfully today. It is having a profound impact on those who embrace it, as it will on those who do not. Over the coming decades it can make a substantial positive contribution to our prosperity.

The six-page submission which follows explains that:

- AI is about people creating machines capable of intelligence. There are eight key AI capabilities.
- Thousands of businesses across all sectors are automating processes with AI to reduce cost and improve services. Much work has begun in the last year, the pace of change is rapid and competitive pressure is high.
- The government could save tens of billions annually through AI-powered automation, whilst improving public services.
- The UK has faced productivity stagnation for the last decade. AI could play a substantial role in reversing that trend.
- AI can also make us happier, healthier and more prosperous in other ways. For example, 1.2 million die on the world's roads every year - the leading cause of preventable death amongst young people. Autonomous vehicles are the ultimate solution.
- AI is a building block in what could amount to a 4th Industrial Revolution. This period could bring extraordinary gains in general prosperity, as in past industrial revolutions.
- The main role of government should be to provide the legal and economic foundations within which AI can thrive. Politicians should encourage and defend entrepreneurship more enthusiastically.
- The political narrative on AI is too pessimistic - on jobs, inequality, monopolies and doomsday scenarios.
- The solution is not to hinder innovation, but to lead from the front, whilst providing sufficient safety to those who are less fortunate. Governments should explore new education and welfare policies, for example, a negative income tax.

About the author

James is a highly commended management consultant, supporting businesses to transform their operations through AI-powered products. He works for

WorkFusion, the market leader in Intelligent Automation software products. He was a founding member of their London office and EMEA expansion.

He previously worked in Strategy & Operations at Deloitte, where he was identified as one of the UK's leading consultants by the Management Consultancies Association. He focused on corporate strategy, innovation initiatives, and transformation programmes.

His clients have included a range of global financial institutions and FTSE 100 companies, as well as the UK government, City of London and Metropolitan Police. James is an associate of the Adam Smith Institute and leads the Archimedes Research Centre. He read philosophy, politics and economics at Oxford, with a particular focus on economic history and international relations.

What is Artificial Intelligence?

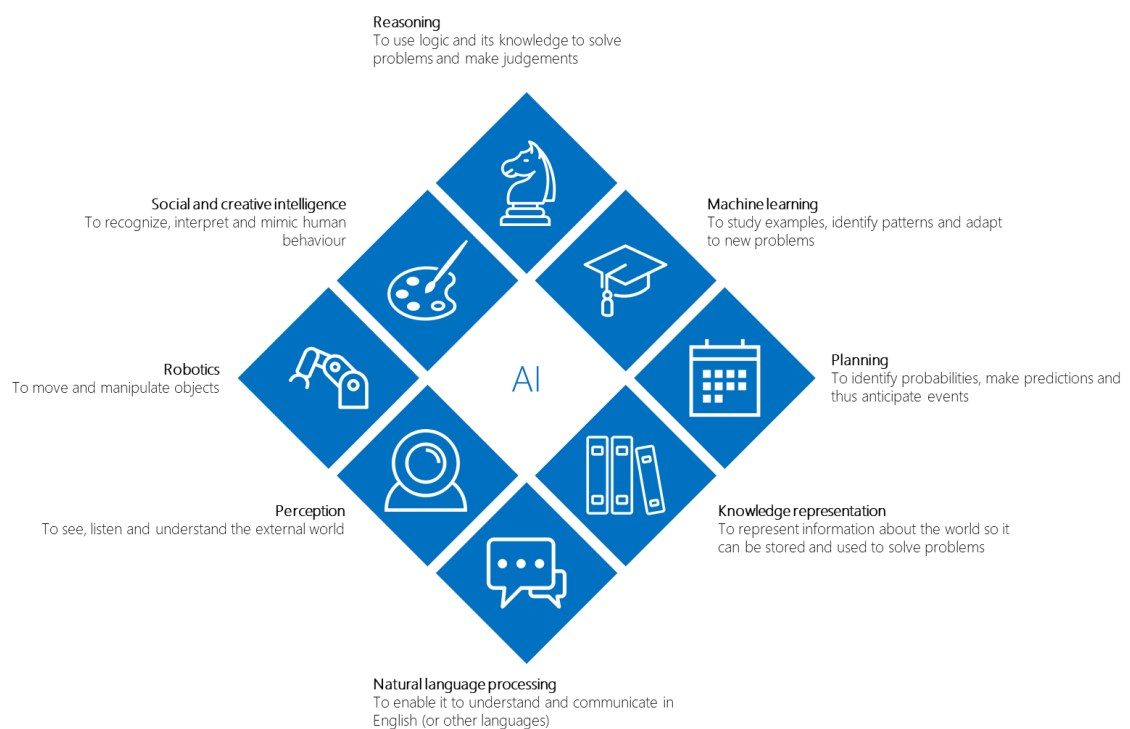
- 1.1 Artificial Intelligence (AI) is a broad field of study, often confused by conflicting usage of terms. To define AI, it is useful to start by considering each word in turn. *Artificial* means made by people. This is in contrast with something that is natural. For example, one might create artificial flowers, that resemble flowers found naturally in gardens. *Intelligence* means the ability to learn, understand and make judgements based on reason.
- 1.2 At its most basic, *Artificial Intelligence is about people creating machines capable of intelligence*. This contrasts with humans and animals which exhibit natural intelligence. This is a useful definition to describe the overall field of study. However, our definition struggles in practice because the *meaning of intelligence*⁷⁸⁷ is open to debate and the *level of intelligence* that needs to be demonstrated is unclear. The intelligence also needs to be *useable* in some practical way to be proven.
- 1.3 Artificial Intelligence traces its roots far back. In the 17th Century, mathematicians like Pascal created some of the first calculators. His machines could add and subtract two numbers. Since then we have made significant advances. Calculators have become routine, rarely considered as Artificial Intelligence.
- 1.4 As we make advances, the scope of AI is disputed – this is sometimes known as the AI effect. "It's part of the history of the field of artificial intelligence that every time somebody figured out how to make a computer do something—play good checkers, solve simple but relatively informal problems—there was chorus of critics to say, "that's not thinking" (McCorduck). Moreover, consumers don't purchase AI, but helpful

⁷⁸⁷ The definition of intelligence and associated concepts like knowledge, learning, understanding, judgement, and rationality are all open to extensive philosophical debate. These cannot hope to be resolved by this paper. For example, Knowledge has been commonly defined as a "justified true belief" since the Enlightenment, and yet this has been challenged since "Is Justified True Belief Knowledge?" (Gettier, 1963). Instead, this paper will focus on examples that would be considered as intelligence in a practical layman or business context without seeking absolute precision.

products and services. The best way to sell AI is make it so easy and beneficial to use, that the recipient need not care about the underlying technology. This reinforces the AI effect.

AI capabilities

- 1.5 To overcome the AI effect in our definition, it is useful to be explicit about the different aspects of intelligence that need to be demonstrated. Alan Turing attempted to do this in the 1950s, hence the famous Turing Test. In this test, an interrogator communicates with a computer and person, and attempts to identify the computer. If the computer is indistinguishable, it wins. To do this successfully, a computer would need various capabilities, including:



- 1.6 At its most advanced, Artificial General Intelligence, machines could demonstrate any intelligence that a human can (as well as or better than humans). This is a valid goal, but does not detract from the advances within and use of, individual AI capabilities. Within individual fields, it is also possible to go beyond average human intelligence – since Deep Blue in 1997, computers have long been the world champions of chess.
- 1.7 We can overcome the AI effect by focusing on the capabilities required for today's AI challenges. To Turing's credit, his test remains relevant to current AI like automating data analysis, and autonomous cars.

Why the hype?

- 1.8 The current AI hype has been driven primarily by wider adoption of machine learning, hardware improvements, and greater investment.
- 1.9 Machine learning⁷⁸⁸ usage is increasingly widespread amongst leading technology companies and exciting new startups. Simultaneously, the provision of complimentary tools has made AI much more useable and beneficial in practice. As Google argued, developing Machine Learning code is just 5% of the challenge, as “the required surrounding infrastructure is vast and complex”.
- 1.10 Computing hardware has become much more powerful, at lower cost. Graphics cards, typically used for video games, have very effective processors for AI applications. Producers claim that recent progress has made it fifty times faster to train an AI neural network. Simultaneously, cheaper servers and cloud-based infrastructure makes it easier to deploy.
- 1.11 Interest in AI has also increased dramatically. This is demonstrable not just in media or political circles, but with businesses and investors willing to spend their money. In 2016, McKinsey estimated that companies invested up to \$39 billion in AI. Private equity firms and venture capitalists invested up to an additional \$8 billion.

Using AI to build a better and more prosperous world

Automating tedious processes

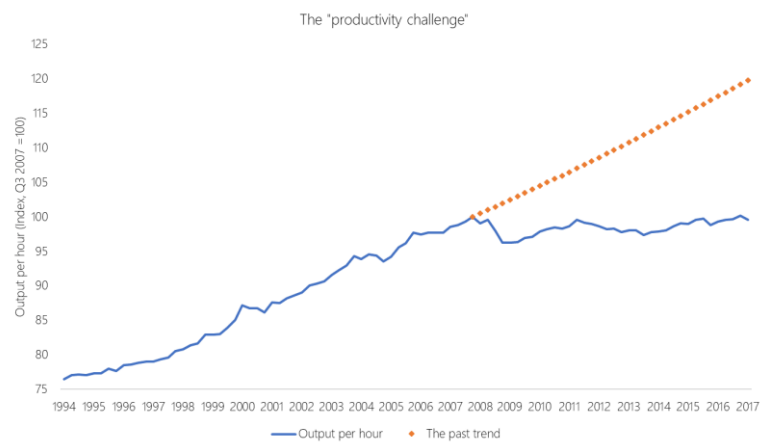
- 2.1 Artificial Intelligence is no longer science fiction for most businesses. Over the last year, businesses across all sectors have started to investigate Artificial Intelligence in some way. Typically, they look for opportunities to automate processes, helping them to improve customer service and reduce cost.
- 2.2 The gateway technology to do this is known as Robotic Process Automation (RPA). To outsiders, this term can be somewhat confusing, as it doesn't involve any robots at all, at least not in the physical sense. It is about 'software robots' performing clerical work, interacting with systems in the same way that people do (with log-ins, passwords, mouse clicks etc.).
- 2.3 These bots are extremely attractive for businesses because they free up staff to work on the more complex parts of the roles, help increase productivity and/or support cost reduction. They also work rapidly, twenty-four hours a day, without making mistakes, and with work closely audited.

⁷⁸⁸ This explains why concepts like machine learning are sometimes colloquially used interchangeably with AI. It has been the biggest area of AI investment and growth.

Implementations are rapid – weeks or months – rather than the years associated with big ICT projects.

- 2.4 Within this niche, a Deloitte survey of executives found that 22% had piloted or implemented RPA in some way and 74% planned to investigate the technology in the next year. This compares against the same survey just one year before, in which nobody had implemented RPA. WorkFusion is a market leading vendor of RPA software. Its data shows that demand is extremely high and that Deloitte's survey is already out of date. Over 4000 business have adopted WorkFusion's RPA in the last year.
- 2.5 RPA's limitation is that it can only automate relatively simple processes – repetitive tasks, clear rules, and structured data. RPA is a gateway to more advanced technologies which use AI capabilities like machine learning. For example, WorkFusion adds 'cognitive' bots (which learn by seeing examples) to eliminate up to 90% of manual work in a process. Standard Bank, Africa's largest bank, has adopted these technologies with enthusiasm. Standard Bank were able to reduce the time it takes to create a bank account from 20 days, to just 5 minutes.
- 2.6 The House of Lords Artificial Intelligence Committee asked about the pace of AI change. For those advising companies and working on AI, it is clear that change is rapid. AI is transforming business today, with quick implementations, and returns on investment.
- 2.7 In our globalised economy, the impetus for businesses to adopt AI is intense. If they don't, their competitors will, or new entrants will disrupt the industry. AI isn't just about automating repetitive processes to help businesses. Leading the first the review of AI in policing, in partnership with a major UK force, it quickly became clear that there were wide applications in government too.
- 2.8 Police Officers must complete a range of laborious paperwork and repetitive processes. This detracts from the time they can spend out on the beat, pursuing criminals and investigating offences. For example, when a force receives intelligence from a partner, this needs to be manually input into their systems. Similarly, the Home Office came under fire in early 2017, as criminal record checks were taking too long to process – frustrating nurses and teachers, and even costing some people their jobs. Here AI could support by rapidly running the necessary queries on different police systems, even if the final reviews are still done by officers. There were also opportunities in areas like the automation of traffic offence reports, licensing, vetting, auditing and even combatting the rise in cybercrime.

2.9 AI could also help other areas of government, from the collection of taxes by HMRC, to the calculation of benefits by DWP and processing of discharged patients in the NHS. With over £2 trillion of state liabilities, the government could save tens of billions annually through AI-powered automation, whilst improving public services.



2.10 More generally, over the last decade the UK has faced what the ONS termed 'productivity puzzle' – i.e. stagnation. AI could play a substantial role in making the UK more productive if government embraces it. AI can also make us happier, healthier and more prosperous in other ways.

Consider our roads

2.11 At least nineteen people died of traffic injuries in the time it takes the average reader to reach this point in the paper. Another is killed every 25 seconds. More than 1.2 million will die on the world's roads every year. It is the leading cause of preventable death amongst young people.

2.12 Even Europe, the safest region in the world sees around 85,000 lives lost annually. The British are amongst the safest drivers in the world, but even we lose more than five people every day. Many more sustain injuries (around 50 million globally), suffering adverse health consequences.

2.13 According to the World Health Organization, 3% of global GDP is lost due to these deaths and injuries. All this happens, despite significant improvements in vehicle safety, the UN's resolution on Road Safety and the fact that it is preventable.

2.14 Artificial Intelligence has the potential to ensure that eventually nobody need die on the roads, except perhaps for those following the antiquated practice of *manual* driving, for leisure or sport. Autonomous, or driverless, vehicles will become the norm within a lifetime.

2.15 For now, autonomous vehicles are still under development, and even when implemented, developing countries will take time to catch up. There are different levels of autonomy. SAE international has created a generally accepted set of definitions for six levels of automotive automation. Tesla cars already come with an autopilot mode, demonstrating 'level 3 -

conditional automation’. This means a Tesla car can take control, but with the expectation that a human driver will intervene when required. The eventual goal is ‘level 5 – full automation’.

- 2.16 Autonomous vehicles should already be safer in most conditions. They have constant 360-degree vision, with no blind spots, as well as ultrasonic sensors and radars. They won’t get distracted, tired, or drink alcohol. They won’t stray from the highway code, or speed.
- 2.17 Early data suggests driverless vehicles are already safer than humans. Yet, human and machine errors mean they’re not 100% safe either. Regulation and the outstanding safety concerns will need to be addressed before autonomous cars become widespread.
- 2.18 So, what is safe enough? Autonomous vehicles don’t need to be perfect to be safer than human drivers and worth embracing. This is of little comfort to potential victims of an accident. However, government should consider the wellbeing of society as whole.
- 2.19 Moreover, with machine learning autonomous vehicles get better the more data they collect. Autonomous car makers are currently training their AI models by shadowing real drivers, allowing the AI to compare and refine its decision making. Advances in driverless cars also contribute to more general road safety, with features like automatic emergency braking, and collision warnings aiding human drivers. Eventually, once autonomous vehicles are dominant, they will communicate with each other to co-ordinate driving, making them even safer.
- 2.20 Autonomous cars would bring a wide range of other benefits, beyond improving road safety. Without drivers, the concept of a car can change to be more pleasant for travellers. It’s likely a typical vehicle will have seats facing each other for conversations, whilst others might have beds or office facilities with full internet access. The average motorist spends hundreds of hours driving each year – time they will instead have for work or leisure.
- 2.21 There will be less pressure on inner cities, as people become more willing to commute longer distances, yet still able to reach work or enjoy evening entertainment. Driverless vehicles will also reduce congestion and emissions by calculating the optimum route and driving approach. Finally, driverless vehicles will improve access and reduce costs. Eventually, few people will own a car, instead paying for journeys per trip, sharing a pool of cars that are utilised more consistently.
- 2.22 The Coalition Government published positive research on “The pathway to driverless cars”. This analysed the legal conditions for testing, producing

and marketing autonomous vehicles in the UK. It concluded that driverless vehicles can legally be tested on public roads in the UK today.

2.23 However, we have made little of the opportunity to become a developmental hub. Testing is allowed, “providing a test driver is present and takes responsibility for the safe operation of the vehicle”. We have yet to proactively legalise ‘level-5’ fully autonomous cars on our roads. Doing so would have the added benefit of encouraging further research and early adoption. Since 2015, autonomous vehicles have fallen off the government agenda. There is also likely to be significant opposition from established industries.

But what about elsewhere?

2.24 Autonomous vehicles are just one example, but they demonstrate the profound impact Artificial Intelligence can have on society. They show that the hype around AI is warranted. The basic technology exists today, and the technology is already being tested on our roads. Without regulatory obstacles, they could start to be deployed more generally.

2.25 Artificial Intelligence is also having a huge impact on the health sector, from helping to diagnose patients, to supporting drug research, and preventing the spread of diseases. In Financial Services, AI is driving highly personalised banking, financial advisory and enhanced trading. Retailers are already predicting customer orders in advance and tailoring adverts – they will soon extend this to personalised products altogether.

2.26 It is notable that the AI revolution is covering all sectors. Even the traditional professions like law and accounting will be forced to change. It has begun with process automation removing mundane work, and will extend to whole functions and services being transformed or replaced.

2.27 There is a wealth of supporting analysis on AI’s huge impact. PWC has argued that Global GDP will be \$15.7 trillion higher by 2030 as a result of AI. It is not a matter of “how likely is AI to develop” but rather “how can we make the most of the revolution”.

What is the role of government?

3.1 AI is a building block in what could amount to a 4th Industrial Revolution. The first started in 18th century Britain as we moved from a farming economy to an industrial and urban powerhouse – with railroad expansion, steam, iron and textile innovations. Economic historians describe the second industrial revolution as the period between 1870 and the world wars, with major new advances like steel, oil, electricity and mass production. Once the world recovered from war, the third revolution emerged with telecommunications, computers and the internet.

3.2 During these periods of substantial economic progress people:

- ✓ Become more productive
- ✓ Solve previously unsolved challenges
- ✓ Receive more goods and services for less
- ✓ Have more fulfilling jobs and more time for leisure
- ✓ See the general population's standard of living rise
- ✓ Live significantly more prosperous lives than those who came before

3.3 This is likely to happen again with the 4th industrial revolution. However, change is challenging, even just considering the impact of AI. Some will lose jobs or need to retrain, even if AI is beneficial overall. At a macro level, some countries will lead the charge, whilst others which are unprepared or try to prevent change, may see stagnation.

3.4 Unsurprisingly, other governments have taken a keen interest in Artificial Intelligence. China has been the boldest, aiming to become the leader, with the industry generating more than \$150 billion by 2030. The Chinese have so far applied for nearly 16,000 AI patents.

3.5 The benefits of Artificial Intelligence will substantially outweigh the costs. The fourth industrial revolution should be embraced. Leaders during these periods of substantial innovation and economic progress see extraordinary gains in general prosperity. AI can transform our businesses and government. It can end our productivity stagnation, address our over-indebtedness and help overcome structural challenges like our ageing population. This is a great opportunity for progress, both for the UK and to help meet the needs of billions in developing countries.

3.6 The main role of government should be to provide the legal and economic foundations within which Artificial Intelligence can thrive. At its most basic, this means maintaining strong property rights, the rule of law, and a flexible economy. Politicians should encourage and defend entrepreneurship more enthusiastically. The cultural and political narrative needs to remain positive or innovation could soon be the enemy.

3.7 Reductions in Corporation tax and schemes that specifically support entrepreneurs like EIS and SEIS are positive, but there is significant opportunity to take these much further. Government should also not be biased to AI alone, as it is part of the 4th industrial revolution, but not the whole.

- 3.8 Governments aren't very good at 'picking winners' in technology, and can be misguided by special interest groups. This is well demonstrated by the recent subsidies in developing countries promoting CFL lightbulbs, only for the market to separately produce a better alternative, LED lightbulbs. So, a flexible approach is preferable to sponsoring preferred companies or setting up a Department for Artificial Intelligence.
- 3.9 For AI projects specifically, the government will need to pass new laws to permit new services and ways of working - driverless cars are an obvious example here. Government will need to hold its nerve when special interest groups campaign against companies that deliver disruptive innovation. Uber is a popular service amongst consumers, but hated by other drivers for the competition it has brought to a previously secure industry. Consider how the established industry will react when driverless cars are available, on a per trip basis, at a fraction of the cost, with greater comfort, safety and reliability.
- 3.10 Governments also need to recognise that AI is complex and requires talent. There is clearly a role for the education system, which is currently relatively poor at preparing the next generation in STEM subjects. Increasing school choice and professional education paths is essential. It is also vital that gifted individuals from across Europe and the wider world can settle in the UK for AI. They should study in our universities, research in our laboratories, support our business and even create their own startups.
- 3.11 At home, the political narrative on AI is too pessimistic. British commentators often focus primarily on the risk to jobs from automation (including for the middle classes), potential rises in inequality, the threat of technology monopolies and headline grabbing doomsday scenarios inspired by dystopian sci-fi films. Government will need to strike a careful balance between fostering progress, and addressing public concerns. The restricted length of submissions only permits some limited considerations on these issues below.
- 3.12 Every industrial revolution has displaced jobs, yet has increased general prosperity. The solution is not to hinder innovation⁷⁸⁹, which is futile in a globalised economy anyway. Instead the UK should lead from the front, whilst providing sufficient flexibility and safety for those who are less fortunate. This can overcome concerns about jobs and equality.
- 3.13 Education will again be important here – everyone needs to keep learning. Artificial Intelligence isn't the first technology to bring automation in the last three centuries, nor will it be the last, and yet the UK has near full employment today. That's because there aren't a fixed number of jobs,

⁷⁸⁹ The French economist Frederic Bastiat reached this conclusion back in 1845. His famous satirical petition by the candlestick makers, who request that government block out the sun to protect their industry from unfair competition, remains relevant today.

they can change, and new ones can be created. 10 years ago, there was no such thing as an app developer, social media manager, cloud engineer or user experience designer, and much fewer elderly patient carers or educational consultants.

- 3.14 Nonetheless, to provide an adequate safety net for those struggling, governments need to also re-evaluate their welfare systems. Policies like a negative income tax⁷⁹⁰ could simplify welfare, whilst providing better support in an era of disruption. With a negative income tax, people earning below a certain amount receive supplementary pay from the government, instead of paying taxes – with those earning nothing guaranteed a basic salary.
- 3.15 As for technology monopolists, there are many reasons to challenge the consensus of pessimism. Firstly, the allure of profits generates the primary incentive for innovation. It is also in the nature of innovation, that companies will often become a leader for a product or service, particularly when it is brand new. Undermining the opportunity to profit from innovation gives entrepreneurs little incentive to risk their time and livelihoods.
- 3.16 Secondly, people generally under invest in the future, and businesses fear risk. Being an entrepreneur is difficult work too. Scale and a track record makes it easier to invest in further innovations, with less concern about financing and greater diversification to minimise risk. Large successful tech companies still have a lot to offer us, in addition to the vast increases in productivity and wellbeing that they have already facilitated through past innovation.
- 3.17 Thirdly, technology monopolies are short-lived unless innovation continues. There are countless examples where competitors emulated, caught up and advanced further. Henry Ford was the undisputed leader in motoring, until the likes of the Dodge Brothers came along with the electric starter. Nokia and Blackberry were leaders in mobile phones, until Apple introduced the iPhone. A technology monopolist can only survive by continuing to lead the innovative pack and is constantly encouraged by the pursuit of challengers. This is the process of creative destruction: “The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers’ goods, the new methods of production or transportation, the new markets, the new forms of industrial organization.” (Schumpeter)
- 3.18 Finally, whilst there may be some special cases (perhaps natural monopolies) the data from the last century shows that the most

⁷⁹⁰ This idea is sometimes credited to 1940s Liberal politician Juliet Rhys-Williams. It is best expounded upon by Nobel Prize winner, Milton Friedman. It is a superior form of a minimum basic income, as it costs less to fund, and is graduated, providing greater incentive for work simultaneously.

longstanding monopolies arise from direct government support or artificial obstacles to innovative competition. Consider telecommunications in the UK before the privatisation of BT. This suggests that governments should avoid artificially creating bad monopolies, in the name of protection or false competition.

As for doomsday scenarios, these threats are typically overstated by those who fear change and by the media who sell content by appealing to our imagination. The UK has much more to lose if it does not embrace the rise of the machines.

References

Boden, Margaret (2006). *Mind As Machine*

Brynjolfsson, Erik and McAfee, Andrew (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*

Brynjolfsson, Erik and McAfee, Andrew (2017). *Machine, Platform, Crowd: Harnessing Our Digital Future*

Deloitte (2016). *The robots are here*,
<https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/Innovation/deloitte-uk-robots-are-here-digital-workforce.pdf>

DfT (2015). *The Pathway to Driverless Cars*,
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/401562/pathway-driverless-cars-summary.pdf

Dreyfus, Hubert (1972). *What Computers Can't Do*

Dreyfus, Hubert (1992). *What Computers Still Can't Do*

Ford, Martin (2016). *Rise of the Robots: Technology and the Threat of a Jobless Future*

Friedman, Milton & Rose (1980). *Free to Choose: A Personal Statement*

Friedman, Milton (2002). *Capitalism and Freedom: Fortieth Anniversary Edition*. University of Chicago Press

Gettier, Edmund (1963). "Is Justified True Belief Knowledge?"

Google (2014). *Hidden Technical Debt in Machine Learning Systems*
<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>

McCorduck, Pamela (2004). *Machines Who Think*

McKinsey (2017). *Artificial Intelligence the Next Digital Frontier?*
<http://www.mckinsey.com/~media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>

Nilsson, Nils (2009). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*

ONS (2017). *Labour productivity Statistical bulletin Jan to Mar 2017*,
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/bulletins/labourproductivity/jantomar2017>

Osborne, Michael and Frey, Carl (2013). *The future of employment: how susceptible are jobs to computerisation?*,
http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

Peoples' Republic of China (2017). *China issues guideline on artificial intelligence development*,
http://english.gov.cn/policies/latest_releases/2017/07/20/content_281475742458322.htm

PWC (2017). *Sizing the prize*,
<http://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>

Raphael, Bertram (1976). *The Thinking Computer*

Russell, Stuart and Norvig, Peter (2016). *Artificial Intelligence: A Modern Approach*

Schumpeter, Joseph (1942). *Capitalism, Socialism and Democracy*

Schwab, Klaus (2017). *The Fourth Industrial Revolution*

Susskind, Richard & Daniel (2017). *The Future of the Professions: How Technology Will Transform the Work of Human Experts*

Turing, Alan (October 1950). "Computing Machinery and Intelligence", *Mind*

WHO (2015). *Global status report on road safety 2015*,
http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/

4 September 2015

Professor Shaun Lawson, Dr Ben Kirman, Dr Conor Linehan and Dr Dan O'Hara –
Written evidence (AIC0127)

**Professor Shaun Lawson, Dr Ben Kirman, Dr Conor
Linehan and Dr Dan O'Hara – Written evidence
(AIC0127)**

[Submission to be found under Dr Dan O'Hara](#)

Professor Mark Lee – Written evidence (AIC0093)

Individual evidence from:

Mark Lee FIET FLSW

Professor of Robotics and Intelligent Systems

1. Artificial Intelligence (AI) has many definitions. I take it to refer to computer software that can perform tasks that are considered to require intelligence. This has broad scope and includes machine learning and intelligent robotics.

2. In 2012 some spectacular results were obtained in a visual recognition competition. The previous record error rate was almost halved by a huge neural network that had to learn 60 million parameters. This breakthrough was made possible by three factors: by using deep networks (that had previously been thought impossible to manage and train), the new availability of vast datasets, and the availability of special hardware (GPU chips). This remarkable success in vision recognition performance was repeated each year since and the error rate is now 2.2%. There is no doubt that this branch of AI (known as Deep Learning) has been responsible for most of the media excitement and the often over-hyped predictions.

3. The Deep Learning success in vision has also led to similar results in speech and language processing, pattern and data analytics, and game-playing tasks. In just a few years it has already become a big bandwagon. For instance, the computer vision conferences report that nearly all the research papers are using deep learning methods. This technology has been a major driver behind the current climate of expectations, and the surge in AI investments and acquisitions.

4. There is a race to own the new technology, with Google as the most active having acquired 11 AI start-up companies since 2012. Apple, Amazon and Ford are a few of the many other examples. It is important to note the source of all this innovation - small teams, often in universities. For example, DeepMind Technologies, was a small British start-up company founded by some academic game developers. They used deep learning to learn how to play 49 games from scratch on a simulated Atari system. After the equivalent of 38 days of game playing, their system could play 29 of the 49 games as well as, or superior to, a professional games tester. DeepMind Technologies was bought by Google in 2014 and is now known as Google DeepMind. In 2015 a deep learning system called AlphaGo from Google DeepMind beat the world champion Go player. Most of the pioneering experts in the field of deep learning have been bought up by the larger companies like Facebook, Microsoft and IBM, where they have easier access to large resources, and bigger salaries. Nearly all these people started in universities and the role of the universities needs to be better recognized, (it is not in the interests of the companies to do this).

5. A major problem with neural networks, and hence all deep learning technology, is that they are impenetrable and inscrutable. There is no way of knowing why a trained network gives a particular result. This means they cannot be used in safety-critical applications. They cannot account for errors, explain their reasoning, or justify their decisions: they are `black boxes'. This rules out many areas such as finance or medicine where even advice needs to be backed up with explanations. Software for safety-critical applications is subject to rigorous Verification and Validation processes; developed by software scientists.

6. In a different approach, IBM (in collaboration with several universities) developed Watson, a computer system specifically developed to answer general knowledge questions posed in natural language. Watson worked by analysing the questions and searching over 200 million pages of data. This was a significant research achievement. The ability to communicate in natural language about the topics covered in an application knowledge base have allowed IBM to find applications for the Watson technology in healthcare, finance, telecommunications, and biotechnology.

7. Hence, transparency is available in some AI technology but not all. Of course, errors can be tolerated in some areas – for example, in conversations, and where estimates and probabilities are involved. Most of the smartphone apps are not expected to have performance guarantees, while train control systems certainly are. It depends upon the risks involved and the acceptability of error.

8. Regarding the ethics of transparency, it is clearly desirable for all systems to be as transparent as possible. Although companies, and governments, argue against it, many, many, problems can be avoided by having access to, and understanding of, the systems that control and influence our lives and fortunes. I know of many issues caused simply by obscurity preventing the correct action or outcome. As an example, self-driving cars are guided by vision systems that are essentially black boxes. However, *in this case*, what we need to know is the error likelihood, not an explanation of how it works. There *will* be crashes caused by the failure of vision systems but the insurance industry will evolve to cover that according the relevant risk estimates. Nevertheless, it is quite different for *driver-less* cars, i.e. with no one on board – in this case the ethics and responsibilities become much more complex, involving the manufacturer and consequentially the AI development company.

9. AI is ubiquitous in modern software. It is often embedded in systems and has become part of the toolkit of modern software engineering. In a sense this means that we should pay attention to the threats and impact of software generally. It is certainly true that issues such as privacy, security, data protection, and transparency are threatened as much by corporate globalization, government policies, and lack of regulation, regardless of AI. Of course, AI is generating a stir, but these matters are urgent, important, and do not need a

special AI focus. I don't want my data to be held by unknown agents, *regardless* of how they obtained it.

10. The adoption of new technology by the public is not straightforward. The major companies think they plan the paths of new innovations but history is littered with products that were rejected by the market. People don't care about technology, only what it will do for them - technology doesn't dictate outcomes. This means that public understanding is crucial. Real engagement is needed to make reasoned decisions and choices, and allow the application of basic common sense to allow constructive influence on the role of technology in our lives (and future lives). This means better information dissemination; but, as noted above, not technological data but information on the purpose and role of the systems and their consequences, significance, and value.

11. Super-intelligence is a theoretical idea that computers could become more intelligent than humans and then evolve to become dominant over humans. This is a diversion. Far more serious threats will materialise before this shows a glimmer of progress. The reason is that all successful AI is *task-based*, i.e. designed for a purpose. Artificial *General* Intelligence (AGI) is needed for *general-purpose* systems and despite years of thinking is still a researcher's dream.

12. Regarding robotics, it is important to distinguish robotics as more than just AI. Putting Watson inside a robot will give a mobile Watson. Much more is needed to obtain intelligent robots that can experience and learn about their environment. Embodiment shapes the structure and meaning of the robot's (or human's) dynamic interaction with the environment, and so this structure captures the totality of the experience gleaned over the developing agent's lifespan. In other words robots will have *subjective* experience and AI is some way behind in addressing this. This has implications for home care and other caring robotic applications; progress is being made but much slower than for AI.

13. Predicting the future.

Regarding the near future, AI growth will continue, with many new developments and applications, particularly using deep learning and big data. Many human level benchmarks will be reached and records broken. Difficult areas are human-machine interaction (empathy, discourse, shared experience) and robotics (real-time subjective learning). The "Deep and Big" approach will superficially solve many of these problems but without verification or explanation facilities they will be barred from safety-critical and sensitive applications. Risk assessment methods are very important in real applications. All engineering and science spend a lot of time and money on risk analysis, and this can be expected to play a big role in the deployment of modern AI.

Regarding futurology, even AI scientists will sometimes lapse into hyperbole, especially if a bit of hype will help their own projects to get more funding. But many (most?) will privately express serious doubts that progress is really as fast

or as simple as the press would like to suggest. There is no news in steady progress; the media likes breakthroughs and excitement. Analysis of predictions of future AI performance shows that they are usually wrong, even when given by experts. Predictions on the implications of technological breakthroughs are just as bad.

5 September 2017

Leverhulme Centre for the Future of Intelligence – Written evidence (AIC0182)

AI: Ethics and Governance

The Issues

Contents

- I.a Introduction: AI, Algorithms and Data
- I.b AI and Data Ethics
- I.c The Challenges of Autonomy
- I.d The Challenges of Intelligence
- II Short-term and Long-term Challenges
- III.a Recommendations
- III.b Conclusion
- Appendix A: Diagram of AI Ethics and Governance Issues

Contact

Dr Stephen Cave

Executive Director (CFI) and Senior Research Fellow
University of Cambridge

On behalf of:

Leverhulme Centre for the Future of Intelligence

AI: Ethics and Governance

The Issues

Written Evidence for the House of Lords Select Committee on Artificial Intelligence

Leverhulme Centre for the Future of Intelligence

1. There is a widespread belief that the rise of Artificial Intelligence (AI) poses both ethical and governance challenges. But what are they? Are they really new? And are they inevitable or more speculative? This paper attempts to give a short overview, showing how the challenges posed by AI relate to those posed by other technologies, and also how the immediate challenges relate to those that might arise in the longer-term.
2. This paper was drafted by members and associates of the Leverhulme Centre for the Future of Intelligence (CFI), a collaboration of the University of Cambridge, the University of Oxford, Imperial College London, and the University of California at Berkeley.

I.a Introduction: AI, Algorithms and Data

3. There is no accepted definition of AI, but the term is often used to describe systems performing tasks that would ordinarily require human (or other biological) brainpower to accomplish (such as making sense of spoken language). There is a wide range of such systems, but broadly speaking they consist of computers running algorithms, often drawing on data. So what makes the ethics of AI systems different from those of the technologies on which they are based, for example, computer ethics or data governance?
4. First, it is important to acknowledge that there is significant overlap between these fields. For example, much of the recent progress in AI has depended upon its ability to exploit large data sets. Where this is the case, many issues in data ethics continue to be relevant. At the same time, there are also distinct challenges posed by AI systems that come from their growing capacities -- i.e., what they are able to do through the combination of increasingly sophisticated algorithms, more data and better hardware. Even if these constituent parts remain of the same kind, AI's increasing abilities will pose new questions (just as the differing abilities of a human baby and an adult pose different moral and legal questions).
5. We could categorise the issues arising from the increased capacities of AI as those arising from a system's *intelligence*, and those arising from a system's ability to make decisions *autonomously*. The diagram at Appendix A maps some of the relations between the challenges arising from these capacities with the challenges arising from use of data.

I.b AI and Data Ethics

6. Much of the recent progress in AI is based on machine learning, by which computers learn to perform certain tasks (e.g., to recognise a cancerous growth) from training on large data sets, then perform these tasks on new data sets. Consequently, many of the worries around data usage are imported into AI, such as the challenges of keeping data **secure**, managing **privacy** and consent⁷⁹¹, or ensuring access to data sets for the **public good**.

7. There are two areas where the combination of AI and personal data raises particular challenges. The first of these is **bias**. Data sets all have limitations -- they have been collected in certain ways, from certain groups at certain times. If a particular system learns from a data set that contains biases, it is likely to reproduce them in its output, such as associating female names with family roles, and male names with careers.⁷⁹² Identifying and correcting such biases poses significant technical challenges that involve not only the data itself, but also what the algorithms are doing with it (for example, they might exacerbate certain biases, or hide them, or even create them)⁷⁹³.

8. One measure that can help in identifying and rectifying bias is ensuring these algorithms are **transparent** -- that is, ensuring it is possible to see not only what data they are using, but also the steps taken in processing it to come to a particular conclusion. For some important machine learning techniques this poses technical challenges, and might involve difficult trade-offs (for example, a more transparent method might be less accurate).

9. Transparency is also an important factor in **interpretability**, which refers to our ability to understand why a system produces a certain output (such as an act, recommendation, or so on). Being able to understand a system in this way is important for many reasons, ranging from being able to give an explanation for a decision to someone affected by it, through to helping to identify a system's limitations or **robustness**. For example, a self-driving vehicle trained on a dataset that is insufficiently varied could malfunction in the real world (such as the car that could not distinguish between the side of a white lorry and the sky⁷⁹⁴).

10. Another area where data ethics intermingle with challenges posed by AI is when these systems are used to **manipulate** people. For example, it was

⁷⁹¹ 'Towards the Science of Security and Privacy in Machine Learning.' Nicolas Papernot, Patrick McDaniel, Arunesh Sinha, Michael Wellman. 11 Nov 2016. arXiv:1611.03814

⁷⁹² 'Semantics derived automatically from language corpora contain human-like biases.' Aylin Caliskan, Joanna J. Bryson, Arvind Narayanan. Science 14 April 2017 : 183-186

⁷⁹³ 'Algorithmic Bias in Autonomous Systems.' David Danks, Alex John London. IJCAI 2017.

⁷⁹⁴ <https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk>

recently demonstrated that insights into a person’s private characteristics can be discerned from their activity on social media.⁷⁹⁵ Drawing on this, sophisticated algorithms could be used to tailor messages to large numbers of individuals to a degree impossible for traditional advertisers. Such systems will increasingly blur the lines between offering, persuading and manipulating.

11. Because of these overlaps with data ethics, and the importance of data in driving the current AI revolution, it is occasionally said that resolving data governance is sufficient to resolve AI governance. But this is a mistake. As the figure in Appendix A shows, although some issues in data ethics are applicable to thinking about AI, there are many other issues that are not related to data, and that have no analogues in data ethics. These are issues arising from an AI system’s distinct capacities, such as autonomy and intelligence, that we will explore below.

⁷⁹⁵ ‘Private traits and attributes are predictable from digital records of human behavior.’ Michal Kosinski, David Stillwell, and Thore Graepel. PNAS 2013 110: 5802-5805. ‘Computer-based personality judgments are more accurate than those made by humans.’ Wu Youyou, Michal Kosinski, and David Stillwell. PNAS 2015 112: 1036-1040.

I.c The Challenges of Autonomy

12. Much of the attraction of AI systems is that they will automate many tasks. In some cases, they will perform tasks simply because we don't want to, perhaps because they are tedious (monthly accounts) or dangerous (bomb defusal). But in other cases, it will be because AI is bringing a distinct advantage, such as performing faster, cheaper or better. We won't realise these benefits if a human is monitoring the system every step of the way -- we will want AI systems to just get on with it (whatever 'it' is). In other words, part of the attraction of AI is its increasing ability to perform tasks **autonomously**.

13. It is this increasing autonomy that gives rise to many of the ethical and governance challenges posed by AI. Take a driverless car: it will need to independently and continually make decisions with potentially life and death consequences (not only the much-discussed but very rare 'trolley problem' cases, but in deciding how aggressively or defensively to drive, for example, or what probability to assign to a child running into the road). It is therefore essential that these decisions are made in ways that align with the values of the relevant stakeholders (the '**value alignment**' challenge).

14. As the decisions made by AI become more complex and consequential, they will also pose difficult questions about moral and legal **accountability**. Complex systems capable of learning might be required to make decisions that could not have been foreseen by the programmers. But where these decisions impact lives -- causing injury, for example -- we will need to know whom to look to for responsibility and redress. This is closely tied to the need to keep systems transparent and interpretable, as discussed above.

15. Some people have argued that some decisions are so important that they should never be made by a machine, no matter how intelligent it is, and that having such decisions automated would violate **human dignity**. Where people draw this line will vary. The case is particularly strong for decisions that are clearly matters of life and death, such as whether to target a certain individual with a lethal weapon and pull the trigger. But there will also be difficult borderline cases, such as AI systems prioritising patients for care.

16. Increasing reliance on autonomous, intelligent systems will also pose new *safety* challenges with ethical and governance elements. One of these is ensuring these systems are robust, as mentioned above. Although systems capable of learning pose new challenges in this regard, nonetheless there is a good deal of established knowledge in testing, verification and standard-setting that can be applied here. More novel is the question of **control**: as machines are given more autonomy, they become less like our ordinary vacuum cleaners and more like our pet dogs. They will become less predictable, choosing unforeseen ways to

achieve the goals we have set, interpreting those goals in unexpected ways, or even developing new goals of their own.⁷⁹⁶

I.d The Challenges of Intelligence

17. This leads us onto issues arising from AI's increasing intelligence. We already have machines that autonomously do things for us, like the thermostat that turns on the heating when the room is cold. But they are mostly so limited in their scope that we would not think to describe them as intelligent. To deserve the name *Artificial Intelligence*, we expect a system to master a task we consider cognitively sophisticated (like beating world-class *Go* players) or a task that involves a broad range of sub-skills and decisions (like driving).

18. While AI systems currently remain narrow in their range of abilities by comparison with a human, the breadth of their capacities is increasing rapidly in ways that will pose new ethical and governance challenges -- as well as, of course, creating new opportunities. Many of these -- challenges and opportunities -- will be related to the impact these new capacities will have on the **economy**, and the **labour market** in particular.

19. Automation has been reshaping the labour market for centuries, prompting some to ask if AI poses a genuinely novel challenge in this regard. Of course, it could anyway be important: many of the world-historical tribulations of the twentieth century would count automation and mechanisation as contributory factors. But at the same time, there is reason to think that AI does transform the challenge, at the very least in heralding an age when machines will be not only stronger than us, but also (in the relevant respects) cleverer. Also, by historical standards, the AI revolution is happening very rapidly: both in terms of the development of the technology and its spread through regions and industries.

20. Previously, many professions have been protected from automation because they require subtle or complex combinations of cognitive (and other) skills. As AI systems increase their capacities, those jobs will also be at risk, including esteemed professions such as medicine and law. It is sometimes said that the focus of AI research could be on *enhancing rather than replacing* humans, but if one AI-enhanced human can do the work previously done by five, then four humans could still become redundant.

21. This gives rise to a range of policy issues. One is how to support those whose jobs become obsolete. This will include not only welfare, but also retraining -- and perhaps finding imaginative new ways to give purpose and dignity to lives in which work plays a much smaller role (bearing in mind we might also be living increasingly longer lives). In addition, the prospect that much AI technology will be held in the hands of the few threatens to exacerbate problems of **social inequality** and immobility.

⁷⁹⁶ 'Concrete Problems in AI Safety.' Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané. 21 Jun 2016. arXiv:1606.06565

22. Where machines are performing tasks for us or alongside us, the combination of their increasing autonomy and intelligence will pose new challenges for our interaction with them. Though these have precedents in current issues of human-machine interactions, they could be taken to a new level by AI. These include the risk that we become overly **dependent** on these systems as a society or as individuals -- such as the driver unprepared for when the car switches to manual, or the doctor who loses the knowledge and skills needed to make a diagnosis or question those made by the machine.

23. Increasing intelligence will also combine with autonomy to exacerbate some of the challenges mentioned above, such as **control** and **value alignment**. It might be obvious that as systems become more powerful and are deployed more widely, it will become ever more important to ensure their decision-making processes reflect the values of the relevant stakeholders in that setting. But as those decisions and settings become more complex and faster-moving this becomes more challenging. Human moral decision-making is highly intuitive and reliant on a mix of abstractions, common-sense and debate. This makes it difficult to program into an AI. But mistakes could be costly, e.g., if that system is running critical infrastructure, or instantiated in thousands of homes or cars.

24. All these challenges will be exacerbated as AI systems become more powerful, and in particular if they approach what is sometimes called **Artificial Superintelligence** (ASI). The term ASI refers to a system that would exceed human capacities across the board. While some commentators believe it unlikely that we will ever develop such a system, the majority of AI researchers believe that we can and will -- eventually.⁷⁹⁷ Certainly there is no reason to think that human-level ability represents any kind of plateau: as with pocket calculators, which are vastly better than humans at arithmetic, once machines can be *as good as us* at a task, it is highly likely that they can also be *better* than us.

25. In addition, high levels of intelligence might bring wholly new questions. We do not know, for example, whether certain levels of intelligence give rise to or require **consciousness**, or other attributes that might lead us to think a system deserves legal or moral **personhood**. But they might. This may seem like a remote prospect, but given the resources currently being invested into AI systems with ever greater capacities, it has never looked more likely -- so we would do well to consider the paths and consequences.

II Short-term and Long-term Challenges

26. Occasionally in discussions of AI ethics, disagreement breaks out between those who believe that talk of conscious machines is a headline-grabbing distraction from immediate challenges like bias and automation, and those who, on the other hand, believe that the potential long-term impact of

⁷⁹⁷ 'When Will AI Exceed Human Performance? Evidence from AI Experts.' Katja Grace, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 30 May 2017. arXiv:1705.08807.

superintelligence completely outweighs any short-term concerns. But there is in fact significant overlap between the shorter and longer-term challenges. Consequently, research directions, institutions and codes of practice developed now could help to address both.

27. A review of the challenges described above, particularly those associated with autonomy and intelligence, suggest that they lie on continua: the challenges grow as the capacities of the system grow. The challenge of managing technological unemployment, for example, will be exacerbated by AI, but also exists now – so measures like supporting adult retraining could reap benefits in the short and long term. Similarly, we need to ensure *now* that decisions made by driverless cars and medical diagnostic tools are aligned with the values of the relevant stakeholders; and by solving these problems, we will be developing the skills to ensure that future more powerful AIs can also be value-aligned.

28. This is not to say that all problems will develop in a linear fashion: it is possible that there will be tipping points -- e.g., a point where labour market disruption tips into major social unrest, or when a system's capacity for self-development enables runaway advances in its abilities. But facing the challenges now will help us not only to prepare for such tipping points, but potentially also to avoid them.

III.a Recommendations

29. There is not space here to explore in depth the potential solutions to all these various challenges. But some **examples of measures** that could help to address a broad range of challenges, in the near and long term, include:

- A. Encouraging professional codes of conduct within the AI industry that reflect the principles of 'ethical design' and 'safe design'. This could extend to the development of safety standards, ethical review boards, and so forth.⁷⁹⁸
- B. Increasing education, not only in computer programming and related skills, but in human-machine interaction, so that citizens are broadly able to assess the capabilities and limitations of AI systems, and work safely alongside them.
- C. Ensuring a broad and diverse range of groups are involved in developing the technology and regulating it, both to avoid building-in bias and to maximise the chance of AI being used for the greater public good.
- D. Ensuring that research and development focussed on increasing the capacity of AI and deploying it in new areas is matched (to a degree) by research into the ethics and impact of this deployment.
- E. Appointing an independent national AI Governance body (that may or may not be the same as any Data Governance body) to analyse short and long-term challenges and make recommendations on their solutions.⁷⁹⁹
- F. Supporting and participating in international efforts to coordinate AI governance.

III.b Conclusion

30. First, this paper aimed to show that while the ethical and governance challenges of AI have significant overlaps with those posed by other technologies, the increasing autonomy and intelligence of these systems will also give rise to new challenges. As these capacities grow, so will the scale of the challenges, for example, in ensuring we do not become overly dependent on these systems, or that we do not lose control of them.

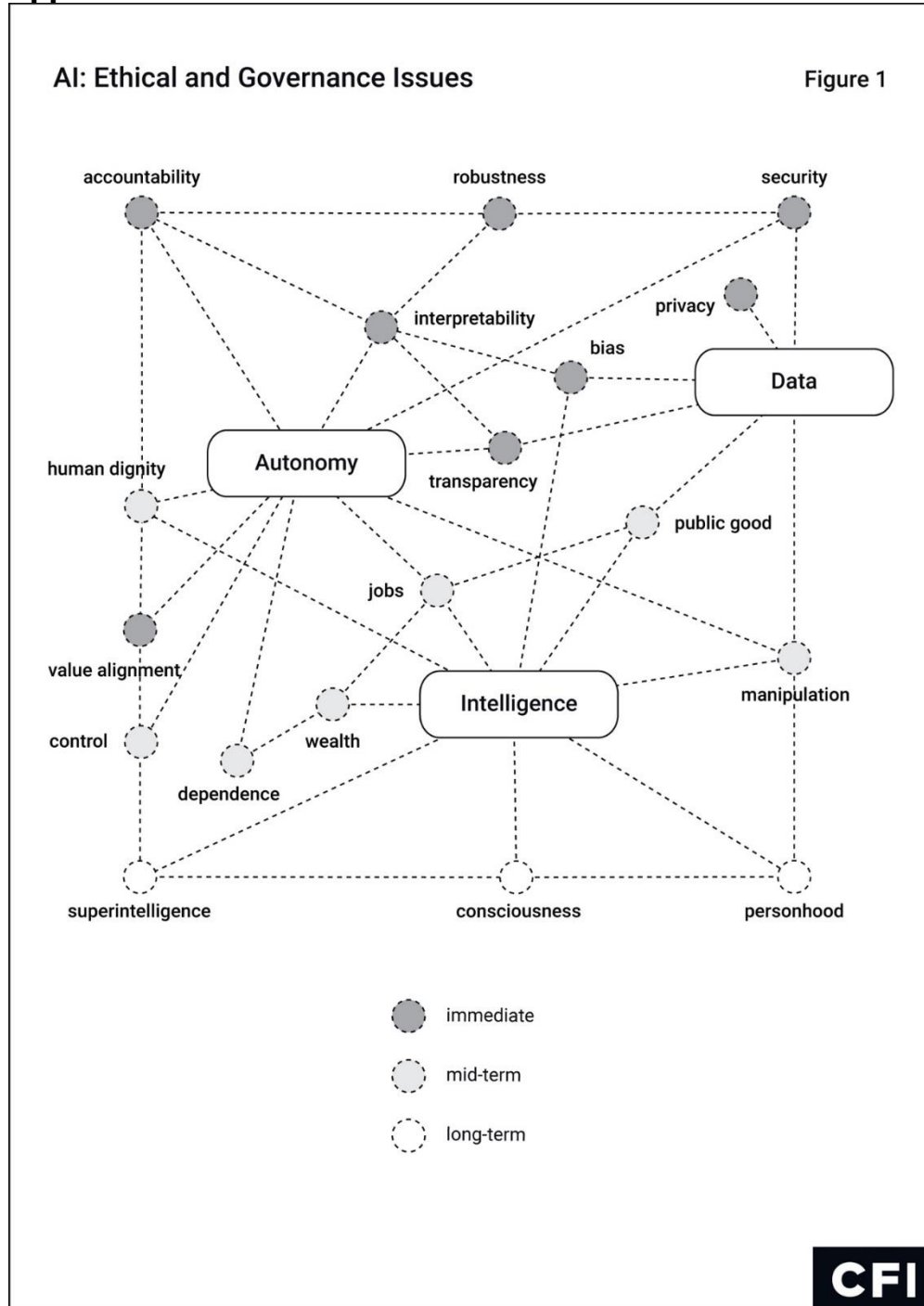
31. Second, this paper aimed to show that there is significant overlap between the challenges posed by AI now, and those it might pose in the future. We do not face a stark choice of focussing on one or the other: rather, we can focus on developing the research capacity, institutional framework, and diverse

⁷⁹⁸ The IEEE's Global AI Ethics Initiative is doing excellent work already on this.

⁷⁹⁹ This was also recommended in CFI (Academic Director, Professor Huw Price)'s written evidence to the House of Commons Science and Technology Committee 'Robotics and artificial intelligence inquiry' (2016), and a recommendation along these lines was subsequently made in the Committee's report on this topic.

community of stakeholders that will help us to address the full range of challenges, and so flourish in the age of intelligent machines.

Appendix A



6 September 2017

Leverhulme Centre for the Future of Intelligence – Supplementary written evidence (AIC0236)

AI & Interdisciplinarity

Written Evidence for the House of Lords Select Committee on Artificial Intelligence

Leverhulme Centre for the Future of Intelligence

I. Introduction

1. This paper is authored by Dr Stephen Cave, Kanta Dihal, Professor José Hernández-Orallo, Dr Seán Ó hÉigeartaigh, and Professor Huw Price at the Leverhulme Centre for the Future of Intelligence (CFI). The CFI is an interdisciplinary research centre that aims to ensure that we humans make the best of the opportunities of artificial intelligence as it develops over coming decades.
2. This evidence shows how research across and above disciplinary boundaries takes place, or should take place, at every stage of the development and introduction of AI, from the creation stage to its functioning in society.
3. Any major new technology raises concerns that are by their nature interdisciplinary – typically, a mix of social, economic, legal, political, and ethical concerns -- which cannot be addressed in isolation from each other, or from knowledge of the enabling technology alone. On top of this, AI brings with it several unique challenges that call for a fundamentally interdisciplinary approach: first, it is a general-purpose technology that will be applied in nearly all domains of life and therefore will be immensely impactful; second, AI systems' autonomy and intelligence raise unprecedented questions that demand new approaches.

II. Interdisciplinarity in the creation of AI

4. AI is itself a product of interdisciplinarity. While rooted in disciplines such as computer science, engineering and mathematics, it is also inspired by neuroscience, psychology and philosophy. For example, DeepMind, founded by neuroscientist Demis Hassabis, draws insights from both cognitive neuroscience and machine learning to develop new breakthroughs in AI with a particular focus on general intelligence and learning, while the Biologically Inspired Robotics Laboratory at the University of Cambridge draws inspiration from physiology and comparative cognition to develop embodied AI.
5. CFI's *Kinds of Intelligence* project explores these relationships in depth, reaching into the fields of psychology, neuroscience, philosophy, computer science, and cognitive robotics. A related ongoing initiative, the *Atlas of*

Intelligence project, aims to map the full range of cognitive capacities that make up 'intelligence', and explore how they have evolved in biological systems and could develop in artificial systems.

III. Interdisciplinarity in understanding AI

6. Understanding AI -- how it works, why it makes particular decisions, how it embeds within wider systems, and so on -- is crucial if it is to be deployed responsibly. This requires technologists to come together with a wide range of scholars from the arts and social sciences. For example: CFI's *Values and Intelligence* project examines how value assumptions and biases can be embedded in AI systems, drawing on philosophy and STS (science and technology studies); CFI's *Trust and Transparency* project draws on this work to develop machine learning algorithms that are as free as possible from these biases, or transparent enough to make them visible; and CFI's *AI Narratives* project examines the way narratives shape the development of the technology, as well as its reception and impact.

IV. Interdisciplinarity in understanding the impacts of AI

7. Different AHSS disciplines provide specific unique perspectives: history gives historical perspective and learning from the past; philosophy provides modes of thought for engaging complex ethical questions; literature, film and media studies are skilled in analysing the functioning and effect of narratives; anthropology draws attention to the webs of interaction that communicate the narratives and the ritualised responses our stories feed into. All of these will be necessary in investigating the ethical and social consequences of AI.

8. Science education will have to adapt in order to accommodate not only knowledge about new technological developments, but also the likelihood that today's children will grow up in a world where many of today's jobs will have changed through automation. Technologists and policy-makers will need to work with education and science communication researchers in order to adapt educational strategies in time.

V. Conclusions and Recommendations

9. Crossing disciplinary boundaries can be difficult for early-career researchers, who are often encouraged to make a name for themselves in a specific discipline. Interdisciplinary research could become more rewarding through:

- A. Supporting high-prestige interdisciplinary forums and journals.
- B. Supporting longer-term career paths, e.g. through interdisciplinary professorships.
- C. Providing interdisciplinary researchers with access to institutional training and support at every stage of their career.

D. Supporting degrees and courses that aim to build expertise on crucial intersections, such as AI and law, AI and policy, or AI and sociology.

10. For interdisciplinary research, it is essential to support, if not mandate, open access. Not only do researchers need access to a very wide range of journals beyond their own fields, but research findings also need to be accessible to policymakers and people in business.

13 December 2017

Leverhulme Centre for the Future of Intelligence – Supplementary written evidence (AIC0238)

AI Narratives

I. Introduction

1. This paper is authored by Dr Stephen Cave, Kanta Dihal, Dr Sarah Dillon and Dr Beth Singler, the Cambridge branch of the AI Narratives project team, at the Leverhulme Centre for the Future of Intelligence (CFI). The AI Narratives Project, joint with the Royal Society, studies the role narratives – understood in the broadest sense of the term – play in the past, present and future development, reception and regulation of AI.

2. We started imagining intelligent machines thousands of years before we could build them. Therefore, as AI and robotics begin to fulfil their promise, they arrive pre-loaded with meaning, sparking associations – and media attention – out of kilter with their capacities. Balancing AI’s potential and its pitfalls requires navigating this web of associations.

II. What can be learnt about the possible development of AI from similarities or differences with previous emerging technologies?

3. Whether the hopes and fears associated with a new technology are managed well can determine whether it is successfully adopted for the public good. Studying the narratives of earlier technologies can provide the knowledge to help situate and inform public debate in anticipation of widespread use of AI. The AI Narratives project held a workshop on this topic at the Royal Society in May 2017, at which leading scientists shared their experiences with the introduction of several emerging technologies.⁸⁰⁰ We are now at the proposal stage for a special issue of the Royal Society journal *Open Science* entitled ‘Narratives of Disruptive Technologies: Lessons for AI’, to be published in autumn 2018.

III. a. What narratives are currently dominating discussions around AI?

4. Literature and other fictional media provide a vast body of thought experiments, or imaginative case studies, about what might happen in an AI future. (We attach a brief history of influential narratives in the Appendix.⁸⁰¹)

⁸⁰⁰ ‘The stories we tell about technology: AI Narratives.’ Susannah Odell and Natasha McCarthy. *In Verba*, 7 December 2017.

<http://blogs.royalsociety.org/in-verba/2017/12/07/the-stories-we-tell-about-technology-ai-narratives/>

⁸⁰¹ This timeline was written by the authors with Professor Elly Truitt and Sankalp Bhatnagar.

Such narratives provide an important data-set for thinking through the social and ethical challenges AI poses.

5. A number of recurrent themes in such narratives dominate discussions around AI, reflecting widespread hopes and fears. These can be seen as a set of dichotomies:

- A. Ease / Obsolescence: there are tales of AI or robot servants enabling humans to have a life of leisure, but at the same time, we fear being made redundant;
- B. Dominance / Subjugation: we pursue AI as a means to attain or demonstrate dominance (e.g., autonomous weapons), but at the same time, fear that our creations will come to dominate over us;
- C. Gratification / Alienation: we create AI to fulfil our desires, e.g., for companionship, but we fear the social isolation or alienation -- the emotional obsolescence -- that could result;
- D. Immortality / Inhumanity: we hope AI and related technologies will give us ever longer lifespans, or even allow us to transcend the body, but at the same time, we fear losing our humanity in the process of transformation.

III. b. Are they the right ones? If not, how should we reframe them?

6. The recurrence of these dominant dichotomies across history confirms their importance in signalling the hopes and concerns that AI raises. These dominant narratives therefore cannot be ignored. At the same time, they pose problems that need to be addressed.

7. Problem A - Perpetuation of polarised responses.

Education on critical engagement with AI narratives can evidence the complexity of thought found therein, depolarising responses and addressing key issues in a more nuanced way.

Recommendations:

- A. Provide schools with free educational material (for use in existing classes) that engages with, but goes beyond, existing narratives around AI.
- B. Support collaborative, interdisciplinary AI research that intersects the Arts and Humanities with STEM disciplines and with government and industry.

8. Problem B - Failure to accurately reflect the range of AI research and development. For instance, dominant narratives revolve primarily around embodied artificial intelligence (humanoid robots).

Recommendations:

- A. Encourage trustworthy, informed and independent science communication from communicators who are able to construct realistic and honest narratives.
- B. Increased development of initiatives to increase the diversity of fictional AI narratives.

9. Problem C - Underestimation of the sophistication of AI Narratives in fiction, and their exploration of the social and ethical implications.

Recommendation:

- A. Ensure participants in the various new bodies being established to consider the ethics, impact and development of AI understand the role and importance of narratives and their study.

10. Problem D - Failure to reflect or encourage equality and diversity. For instance, the gendered and racialised characteristics of robots both determine how they are treated by humans and how the groups represented by these characteristics continue to be treated.

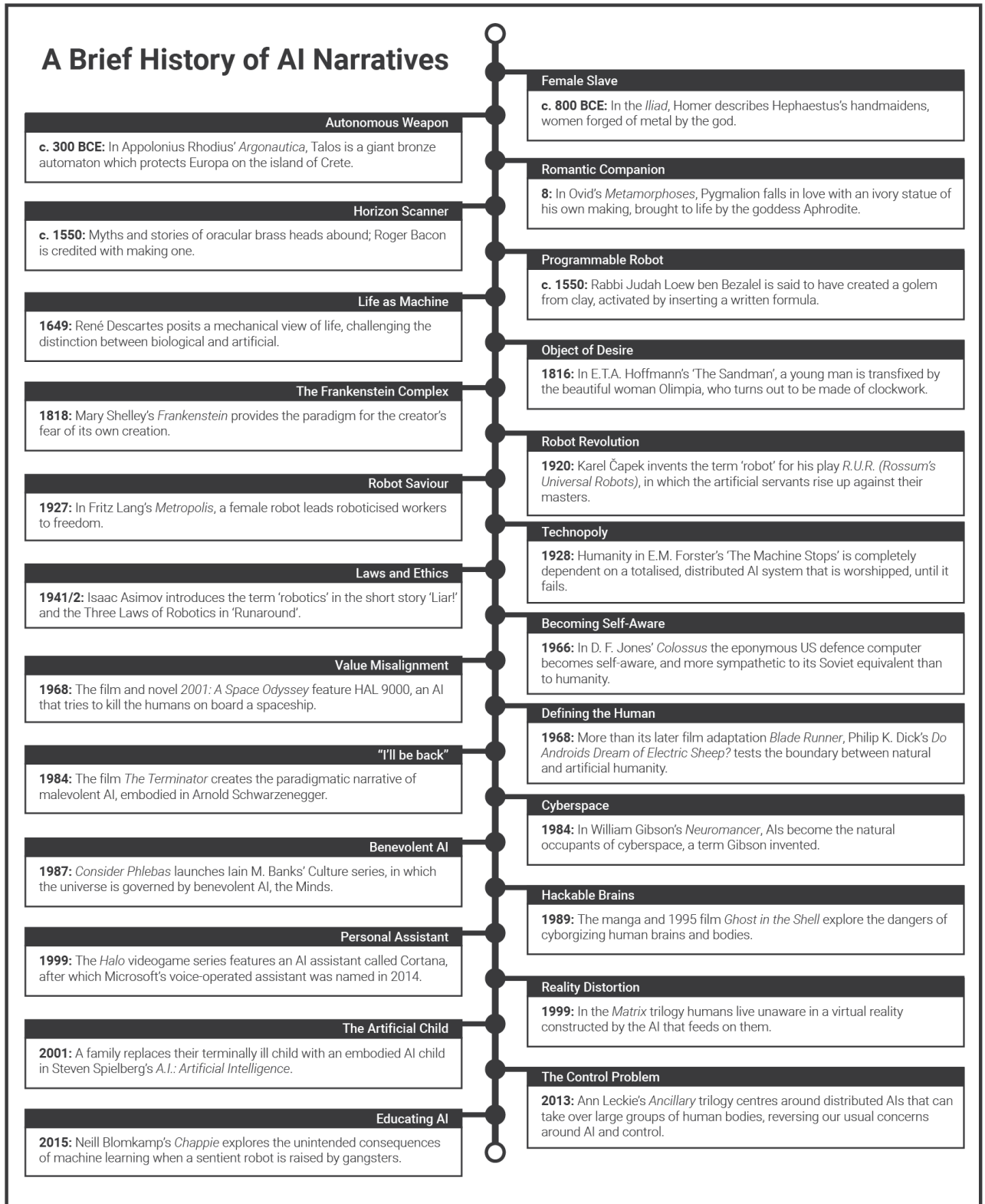
Recommendations:

- A. Facilitate the development of diverse narratives that take into account underrepresented voices, so that AI development pursues the best possible outcomes for all of society.
- B. Facilitate and encourage ongoing public dialogue to make sure that AI develops according to the various needs of society.

11. The AI Narratives project is currently planning the following interventions:

- A. A systematic survey of the history of AI narratives.
- B. Events in collaboration with the Royal Society to increase the diversity of AI narratives.
- C. A presence at existing high-level AI conferences to engage those developing the technology on how narratives form an integral part of AI research and development.
- D. Supporting a four-part short documentary film series on AI, made by Dr Beth Singler at the Faraday Institute for Science and Religion, which will be disseminated for free along with educational material for schools in 2018/2019.
- E. A workshop at the Royal Society in 2018 to review insights on supporting well-founded and diverse AI debates.

- F. A sub-project, What AI Researchers Read, supported by the Royal Society and led by Dr Sarah Dillon, which investigates the influence of imaginative literature on AI researchers' thought and practice.



Leverhulme Centre for the Future of Intelligence and Centre for the Study of Existential Risk – Written evidence (AIC0237)

Horizon-scanning and foresight in artificial intelligence

Seán Ó hÉigearthaigh and José Hernandez-Orallo

Artificial intelligence is set to impact every aspect of life. Progress in recent years has crossed a threshold whereby advances in the science of AI rapidly result in applications with societal impact, in turn encouraging further investment into AI research. With developments occurring ever more rapidly, and with ever more far-reaching consequences, it is useful to draw on a range of methodologies to inform our thinking on:

- a) The expected capabilities of AI systems likely to be developed on the foreseeable horizon - what they will be able (and will not be able!) to do, what resources they will require, the circumstances in which they are likely to be useful
- b) The expected impacts that these systems will have when deployed in various real-world settings - their implications, for example, for scientific development, automation of employment-relevant tasks, and physical and cybersecurity.
- c) Societal, ethical and legal challenges that are likely to be raised by such developments

Prediction is fraught with error, particularly when looking more than a couple of years ahead. However, value can be gained by narrowing down the range of possibilities. It is also often the case that unanticipated potential consequences of particular developments can be identified quite quickly and clearly by bringing together experts from fields likely to be affected by artificial intelligence, and providing them with a setting in which to discuss cutting-edge progress in AI with field leaders.

CFI and CSER are exploring a range of techniques for forecasting and preparing for future impacts of artificial intelligence. These include:

(1) Interdisciplinary workshops

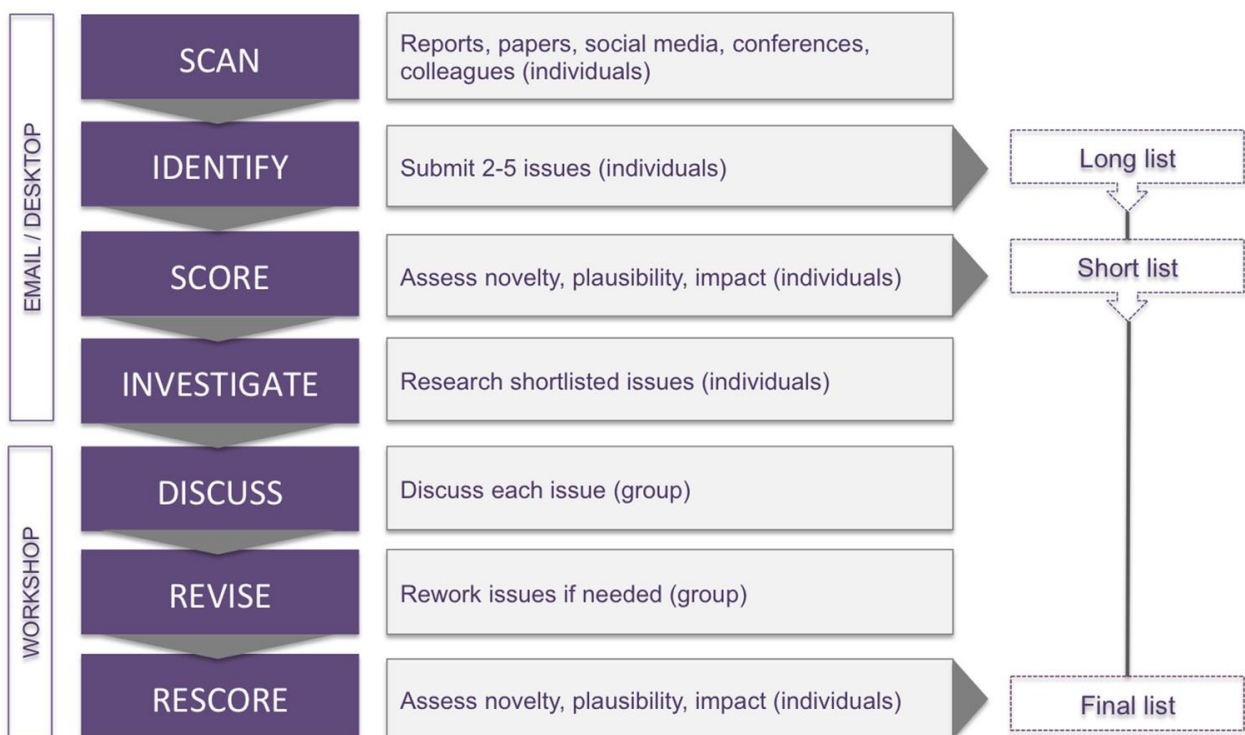
Example 1: 'Malicious use of AI' workshop (February 2017; report forthcoming in January 2018). CSER, CFI and FHI brought together research leaders in machine learning alongside experts in cybersecurity, physical security, and political science to analyse the potential impacts of artificial intelligence on these latter domains. This workshop identified a number of key challenges (see 'malicious use of AI' submission), as well as broader trends relating to changing dynamics in cyberattack versus cyberdefense, information manipulation, and growing vulnerabilities in existing physical systems (e.g. infrastructure) and related

emerging technologies (e.g. drones and household robots) that require further analysis.

Example 2: ‘Data Analytics for Sustainability and Environmental Risk’. CSER and collaborators brought together machine learning experts with experts in climate science, biodiversity loss and sustainability to identify research problems that recent advances in data analytics and machine learning could fruitfully be applied to. One example identified and discussed was the analysis of patterns of melting and reforming in arctic sea ice, where improvement in analysis could lead to more accurate climate prediction. See <https://www.cser.ac.uk/events/daser/>

(2) Delphi-style expert elicitation

CSER also uses more structured expert elicitation techniques, including a modified version of the DELPHI technique (<http://onlinelibrary.wiley.com/doi/10.1111/2041-210X.12387/abstract>). This was recently used to identify a set of emerging and under-recognised issues relating to biological engineering that were deemed to have potentially globally significant impacts on (i) 5 year (ii) 5-10 year (iii) >10 year time horizon. The technique is illustrated by the graphic below:



In brief, 27 experts from biological engineering and related fields were recruited. They generated 70 issues that were independently assessed by other members of the group for scientific plausibility, global impact, and lack of recognition outside of biological engineering. A shortlisted set was then discussed and

reworked where necessary as part of a workshop, then rescored, providing a final list of 20 emerging issues. These are published in the following paper: <https://elifesciences.org/articles/30247>.

The approach shows good promise for application to identifying under-recognised issues relating to the impacts of AI.

(3) Scenario red/blue-teaming

One useful approach for mitigating risks or negative uses of a technology involves experts developing plausible scenarios, and using these scenarios as a basis for a red/blue team exercise, in which one team takes the role of an 'attacker' and one the role of a 'defender', identifying workable solutions to potential harmful developments. A workshop that employed this approach was held in February 2017 by CSER's Jaan Tallinn, Microsoft's Eric Horvitz, and the Bulletin of Atomic Scientists' Lawrence Krauss. Several scenarios - on adverse behaviours of reinforcement learning agents, and AI-enabled cyberattack - were provided by CSER researchers; other scenarios centred on combating the spread of false information and propaganda, and AI-enabled manipulation of stock markets. A full report is expected in early 2018.

(4) Ongoing initiatives: Milestones, measurement and AI footprints

We have several additional early-stage initiatives around AI forecasting and measurement. The AI Milestones initiative aims to identify benchmarks that can be used to clearly identify and characterise fundamental breakthroughs in the capabilities of AI systems. This initiative is motivated by a series of recent breakthroughs in particular tasks (such as game playing or video tagging). While these breakthroughs have been widely covered in the media, they sometimes do not represent the most significant breakthroughs in AI progress; sometimes more fundamental achievements have gone mostly unnoticed.

It is therefore necessary to design both a series of benchmarks that can assess what AI is able to do today, and a battery of challenges that will test for, and motivate, future progress. The benchmarks will be oriented towards the evaluation of fundamental capabilities of AI systems that can then find application in a range of settings, rather than incremental performance improvements on a specific, narrowly-defined task (there already exist a growing number of well-defined benchmarks for the latter, such as performance on image classification datasets, language translation, etc).

Relatedly, in considering the significance of a new AI system or technique, we should consider not only the performance of the new system, but also all the resources that are required to apply the new technology in various settings. These resources may include: the configuration of the task, the amounts of data required, the extent of data formatting and labelling required by human users, computational resources needed, testing, and programming work needed to integrate and deploy the system in different contexts. We aim to develop a series

of metrics, which we refer to as “AI footprints” to provide an estimate of these factors. A low ‘AI footprint’, for example, would indicate that an AI system is likely to be useful for a wide range of users on a wide range of problems (e.g. due to a low need for computational resources, data, or reprogramming effort).

(5) FHI - tracking and projecting progress in hardware and algorithms

Our collaborators at the Future of Humanity Institute have sponsored valuable work to track, model and predict progress in hardware and algorithmic performance metrics with relevance to AI; much of this work can be accessed here: <https://aiimpacts.org/>

FHI researchers have also surveyed experts in AI to gauge their views on when AI capabilities may achieve human-level performance in a range of domains: <https://arxiv.org/abs/1705.08807>

(6) Work elsewhere:

Some other work has focused on the analysis of objective criteria, such as task performance, research investment or bibliographic indicators. The EFF metrics (<https://www.eff.org/ai/metrics>) is the most thorough source of performance results for a wide range of AI tasks, in some cases covering more than a decade, including contributions from our colleagues at the FHI. It is nonetheless difficult to make extrapolations from this data, which in many cases has a logistic shape, with the steepest increase around 2015-2016 and some slower increase afterwards.

A more general set of indicators, mostly meant for the media and policy makers, is the AI index (<https://aiindex.org/>), developed by the Stanford 100 Year Study on AI. The report provides a range of plots and summarised data on the volume of activity in academia and industry, public interest in AI, technical performance (a simplified version of the EFF metrics above) and derivative metrics such as the AI vibrancy index. The report also includes some analytical insight from leading researchers in the field.

A very comprehensive report on the effect of computerisation (not limited to AI) on employment (<http://www.sciencedirect.com/science/article/pii/S0040162516302244>) was published by Frey and Osborne (CFI associate), estimating the probability of automation for 702 occupations.

Luke Muehlhauser published a detailed analysis of lessons from past AI forecasts (<https://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/what-should-we-learn-past-ai-forecasts>). This is useful for comparison and contrast between the current AI ‘boom’ and past ‘AI summers’ (1960s and 1980s) where expectations for, and investment in, artificial intelligence were particularly high.

Recommendations:

- In order to remain in a strong position to guide and respond to advances in AI, the UK government should hold, sponsor and draw on a range of progress-tracking and foresight programmes both on AI capabilities, and on the impacts of these developments on related fields and on a range of real-world contexts.
- These exercises should be regularly updated or re-run.
- The exercises should not be limited to AI researchers, but should involve experts in law, economics, policy, social science, risk and other fields so as to better anticipate the broad-ranging impacts of AI.
- An appropriate government partner may be the Government Office for Science's Horizon-scanning division

13 December 2017

Leverhulme Centre for the Future of Intelligence and Centre for the Study of Existential Risk – Supplementary written evidence (AIC0239)

Long-term Catastrophic Risk from Artificial Intelligence

Haydn Belfield and Seán Ó hÉigearthaigh

1. This submission primarily addresses the question (vii), though we note that it is relevant to all the questions:

(vii) What are your views on the possible existential threats that AI pose to humanity, and human agency? Do we need to discuss these, or are they a distraction from more mundane issues?

Introduction

2. The field of AI is advancing rapidly. Recent years have seen dramatic breakthroughs in image and speech recognition, autonomous robotics, and game playing. This is leading to ever-increasing scientific interest, and governmental and commercial investment, in AI, which is very likely to support continued progress. The benefits of this progress are tremendous: new scientific discoveries, cheaper and better goods and services, medical advances represent but a few. It also raises near-term concerns: privacy, bias, inequality, safety and security. But a growing body of experts within and outside the field of AI has raised concerns that future developments may pose long-term, high impact safety and security risks.

3. Most current AI systems are 'narrow' applications – specifically designed to tackle a well-specified problem in one domain, such as playing a particular game, or classifying images. Such approaches cannot adapt to new or broader challenges without significant redesign. While the system may be far superior to human performance in one domain, it is not superior in other domains. However, a long-held goal in the field has been the development of artificial intelligence that can learn and adapt to a very broad range of challenges while operating in a wide range of environments. Recent progress has been encouraging: for one example, a variant of DeepMind's AlphaGo was able to learn to outperform both human experts and game-specific algorithms in Go, Shogi and chess without having been specifically designed for any one of these games⁸⁰². This system was provided no domain knowledge other than the rules of the game in question, and achieved these performance levels after several hours of playing itself. There is of course a huge gulf between an algorithm capable of learning multiple board games, and a system that approaches the level of general problem-solving ability that a human has. However, there are likely to be continued advances in

⁸⁰² <https://arxiv.org/pdf/1712.01815.pdf>

research on systems that 'learn to learn' without being hand-crafted for a particular challenge, and many practical and scientific applications for such flexible, adaptable systems.

Transformative Artificial Intelligence

4. As AI systems become more powerful and more general they may at a future point achieve performance superior to human capability in many or nearly all domains. While this might sound like science fiction, many research leaders believe it possible⁸⁰³. Were it possible, it might be as transformative economically, socially, and politically as the Industrial Revolution. This could lead to extremely positive developments, but could also potentially pose existential risks from accidents (safety) or misuse (security).

5. On safety: our current systems often go wrong in unpredictable ways. There are a number of difficult technical problems related to the design of accident-free artificial-intelligence. Aligning current systems' behaviour with our goals has proved difficult, and has resulted in unpredictable negative outcomes. Accidents caused by a far more powerful system would be far more destructive.

6. Two arguments about transformative AI have been influential. The 'orthogonality thesis' states that intelligence and final goals are independent – any level of intelligence could be combined with any final goal. If an AI system is very intelligent, this intelligence does not guarantee that its final goals necessarily will benefit humanity. The 'instrumental convergence' thesis states that whatever final goal an AI system has, a number of instrumental goals – prevent 'aggressors' from turning it off, acquire more resources – are likely to emerge as part of a strategy to achieve its final goals. A powerful but poorly aligned AI system that takes actions in the world to achieve those instrumental goals could produce catastrophic consequences.

7. On security: transformative AI would be an economic and military asset to its possessor, perhaps even giving it a decisive strategic advantage over other actors. Were it in the hands of bad actors, they might use that advantage in harmful ways. If two or more groups competed to develop it first and thereby gain that advantage, it might have the destabilising dynamics of an arms race.

Current work

8. There is great uncertainty and expert disagreement over development timelines for transformative AI. However, even in the face of this uncertainty, there is valuable work that can be done now, both on technical design and broader question of strategy, governance and responsible development. Much of this work will have relevance to near-term issues, but will also set the

⁸⁰³ For example, see a recent survey of AI research leaders: <https://arxiv.org/pdf/1705.08807.pdf>

foundations for addressing the challenges posed by more powerful future systems.

On technical AI safety research:

9. Many of the longer-term catastrophic concerns relating to loss of control of AI systems, unpredictability of actions and strategies pursued by AI systems, and the difficulty of designing well-specified goals, guidelines and values for AI systems relate to fundamental issues that we can begin to explore in the design of current day systems. For example:

- 'Concrete problems in AI safety' lays out set of fundamental problems relating to unanticipated and unwanted behaviours of reinforcement learning agents and machine learning systems; these have relevance to near- and longer-term AI systems.
(<https://arxiv.org/pdf/1606.06565.pdf>)
- The Future of Humanity Institute, a partner of the CFI, has collaborated with DeepMind to explore from fundamental principles the design of autonomous AI agents that are 'safely interruptible' – i.e. they will not seek to avoid or subvert interruptions to their performance if a human operator feels it is necessary to shut the system down. Modern-day AI systems are not sufficiently advanced for this to be a cause for concern, but such work will be valuable in the fundamental design of more general and autonomous future systems.
(<https://intelligence.org/files/Interruptibility.pdf>)
- CFI partner CHAI (Centre for Human Compatible AI) is exploring methods for AI systems to infer goals and values from observing the behaviour of humans (cooperative inverse reinforcement learning), rather than being provided hard-coded goals and specifications.
(<http://humancompatible.ai/publications>)
- DeepMind are developing environments in which to test AI agents for safety-relevant behaviours – these include "safe interruptibility, avoiding side effects, absent supervisor, reward gaming, safe exploration, as well as robustness to self-modification, distributional shift, and adversaries"
(<https://deepmind.com/research/publications/ai-safety-gridworlds/>)
- Research to make AI systems more transparent and interpretable is likely to aid in developing future AI safely, as such research will allow the underlying function of AI systems to be easier to monitor and predict, and the impacts of innovations in AI research will be more straightforward to anticipate (<https://nips.cc/Conferences/2017/Schedule?showEvent=8795>).
- A burgeoning area of research is 'reliable machine learning in the wild': the design of AI systems so that the systems will perform reliably, or will provide clear indicators when they cannot perform reliably, in environments and contexts for which their training is insufficient. While

this is relevant to present-day systems (e.g. self-driving cars in unusual weather conditions), fundamental research in this area will lay important foundations for the safe design of more powerful systems – systems which will have a greater range of actions available to them, and which will be performing in a wider range of real-world environments (<https://sites.google.com/site/wildml2017icml/>)

10. It is of course the case that not all aspects of the safe design of future transformative AI systems can be worked on in the context of more primitive, modern-day systems, and ongoing research on more fundamental long-term issues is also necessary (drawing on philosophy and fundamental principles from computer science). However, such work can happen alongside, draw on, and not distract from research on near-term challenges.

On broader responsible development of AI, and avoidance of long-term risk:

11. Given the uncertainty over the timelines involved in the development of transformative AI systems, many of the most valuable steps for the safe and beneficial development of AI are more general, and have relevance to both near- and long-term development of AI. These include:

- Encouraging norms of cooperation and collaboration around the safe development and beneficial application of AI between research groups and companies.
- Encouraging broader stakeholder engagement around the beneficial development and deployment of AI, with experts from disciplines such as law, risk, security, policy, economics, social sciences and global development working alongside AI research leaders to anticipate and guide the development of AI. Participation from a range of the communities who will most be affected by AI in the coming years is also important.
- Supporting programmes to monitor trajectories of progress in AI, and to predict and analyse the impacts that particular advances in AI are likely to enable (see 'horizon-scanning and forecasting' submission).
- Encouraging a greater level of engagement around the development of AI on a global level. In particular, developing a greater level of active dialog between research leaders and policymakers in the West (US/Canada/UK/Europe) and the East (China, Japan and India in particular).

There are promising steps being taken on all of these priorities, and UK companies, academic institutions and policy bodies are playing a key role.

Conclusion:

12. At present, we broadly support the following statement from the White House OSTP's 2016 report on the Future of Artificial Intelligence (from its section on long-term concerns about super-intelligent General AI):

“The best way to build capacity for addressing the longer-term speculative risks is to attack the less extreme risks already seen today, such as current security, privacy, and safety risks, while investing in research on longer-term capabilities and how their challenges might be managed. Additionally, as research and applications in the field continue to mature, practitioners of AI in government and business should approach advances with appropriate consideration of the long-term societal and ethical questions – in addition to just the technical questions – that such advances portend.”⁸⁰⁴

13. Transformative artificial intelligence could be possible within the coming decades. It could have negative as well as positive consequences. There are useful research directions that can be pursued now that have relevance to the performance and challenges raised by AI systems in the near term, but are also likely to lay the foundations for work relevant to the longer-term challenges posed by future AI systems. Our view, therefore, is that such work is worth supporting, and that some research effort should be dedicated to the longer-term possibilities that artificial intelligence raises as well as the nearer-term issues.

13 December 2017

⁸⁰⁴

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

LexisNexis UK – Written evidence (AIC0164)

6th September 2017

Who we are

LexisNexis UK is a leading provider of content and technology solutions. We are part of the RELX Group, a FTSE 30 UK-based provider of information and analytics for professional and business customers across a range of industries.

We make our comments drawing on our expertise in developing software platforms and tools that enable professionals in legal, corporate, tax, government, academic and not-for-profit organisations to make informed decisions and achieve better business outcomes.

We have focused our comments, though not exclusively, on the impact of artificial intelligence technology in the legal industry.

LexisNexis UK considers artificial intelligence to be any system capable of performing tasks utilising some aspects of human intelligence such as logic, reasoning, learning and deduction. In our global business, we are investing in artificial intelligence to help benefit the legal industry, including in the following areas:

- **Assisted decision making:** Lex Machina⁸⁰⁵, our legal analytics platform, mines litigation data in the US to help attorneys prepare for litigation based on data trends.
- **Automated review:** Our technology scans legal documentation⁸⁰⁶ to review and optimise documentation through best-practice clauses, enhanced drafting and case citation checking.
- **Natural language research:** Lexis Answers⁸⁰⁷ utilises machine learning and natural language processing to make legal research easier to use and more efficient.
- **Analytical research:** Ravel Law⁸⁰⁸ utilises machine learning to provide legal research and insight from massive amounts of legal data.

As a data business, our parent company has also developed tools to help decision-making using big data. For example, the HPCC systems platform⁸⁰⁹ is a hugely powerful computing system that uses machine learning to extract insight from data.

⁸⁰⁵ <https://lexmachina.com/>

⁸⁰⁶ <http://www.lexisnexis.co.uk/en-uk/products/lexisdraft.page>

⁸⁰⁷ <https://www.lexisnexis.com/infopro/keeping-current/b/weblog/archive/2017/06/29/you-ask-lexis-174-answers-new-machine-learning-feature-on-lexis-advance.aspx>

⁸⁰⁸ <http://ravellaw.com/>

⁸⁰⁹ <https://hpccsystems.com/about>

We believe that this technology has great potential for an increasingly data-driven legal industry.

Executive summary

- The legal sector is already recognising the potential for artificial intelligence to introduce efficiencies, widen access to justice and help create a sustainable and successful industry.
- The primary risks are the potential for loss of jobs, bias and discrimination in automated decision taking and unregulated self-service pathways.
- The evolving landscape of artificial intelligence enables new types of skills and roles to enter the legal profession. This is already manifesting itself through the employment (both in our business and in the wider industry) of data scientists and knowledge engineers working together with legal teams.
- The approach of the Government should be to support research and innovation in the legal industry and beyond while at the same time undertaking further research and consultation through the formation of appropriate independent bodies.
- Any proposed regulation of artificial intelligence must recognise that a one-size-fits-all approach will not be appropriate. Regulation should be risk-based and focus on the outcomes of artificial intelligence technology rather than the technology itself.

Responses

Question 6: What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

1. Sectors that generate or have access to large sets of data, including the legal sector, stand to gain the most from advances in artificial intelligence. One of the key benefits of the technology is a reduction in the need for human analysis of large sets of data, and the creation of new uses for that data through programming and analysis not currently possible for humans.
2. The legal artificial intelligence industry still remains in a nascent stage but the technology is beginning to change the way lawyers work and the way clients (both business and consumer) access legal services. For businesses and law firms, analytical software that uses artificial intelligence techniques such as natural language programming has been available in the market for a number of years⁸¹⁰, but research and development of new products is increasing rapidly.

⁸¹⁰ Lex Machina, now a LexisNexis group company, was launched in 2009.

3. Technology is making it easier for new entrants to disrupt the market. The widely-publicised DoNotPay bot, for example, has reportedly⁸¹¹ helped motorists overturn 160,000 parking fines and is now providing assistance with asylum claims.
4. Although the use of artificial intelligence is not yet widely prevalent in the legal sector, it is expected to become more so in the next five to ten years as purchasers of legal services become more demanding of cost-savings driven by firms' utilisation of technology.
5. In the short to medium-term, the following uses of artificial intelligence in the legal sector will increase the performance, sustainability and efficiency of legal service providers:
 - a. Document review and automation;
 - b. Online self-service pathways (so-called 'robo-law');
 - c. Risk analysis;
 - d. Research;
 - e. Judgment analytics and decision-making; and
 - f. Contract and case management.
6. This move to a technology-enabled legal sector must be encouraged. The potential benefits are extensive and range from enhanced access to justice for consumers (including those unable to afford legal fees), to the long-term sustainability of a legal system that is globally recognised and respected.
7. Concerns have been raised about wide-spread use of technology leading to significant loss of jobs in the sector. These concerns are considered in the paragraphs below.
8. Technology has been changing the employment landscape in UK law firms for decades. The use of digital dictation and case management software has led to the loss (or offshoring) of support roles such as legal secretarial and administration, for example.
9. There is evidence pointing to a sharp increase in the pace of this change in the next few years. A report⁸¹² published by Deloitte in February 2016 estimates that a technological tipping point might occur around 2020. It is predicted that organisations that do not foresee this major change, and address it, will become unsustainable and fail.

⁸¹¹ <https://www.theguardian.com/technology/2017/mar/06/chatbot-donotpay-refugees-claim-asylum-legal-aid> (accessed 6 September 2017).

⁸¹² [Developing legal talent, stepping into the future law firm](#). February 2016 (accessed 6th September 2017).

10. Loss of jobs in the legal sector due to technology is likely to continue to be concentrated in administrative roles (as it has been for some time), but it is possible that the rise in use of artificial intelligence will impact legal roles at the junior end of the scale including paralegals, legal executives and junior lawyers. These employees are commonly utilised to complete high-volume tasks that are routine but still require analysis and reasoning⁸¹³.
11. However, when artificial intelligence can undertake these tasks reliably, law firms may be able to refocus the work of junior lawyers away from routine tasks to tasks that can add more value and better utilise the lawyer's abilities. This might include hybrid legal and technology tasks such as supervising the artificial intelligence program or verifying its results.
12. One of a lawyer's main skills is problem-solving. This potential shift for junior lawyers away from mundane and routine work will allow the lawyer, as one LexisNexis UK customer has put it, to 'trade up to a better class of problem'.
13. The correct response to the impact of AI is not to slow down or over-regulate the development of technology in the legal sector. It is critical that technological development is encouraged to maintain the sustainability and competitiveness of the sector and to ensure that talented people are not lost to other industries.
14. A rethink is required of how we can equip people with skills to enable them to understand and use legal technology and to actively shape the transition that the sector is going through. This may include compulsory technology modules (covering subjects such as data analytics) as part of solicitor apprenticeships, the legal practice course, and the bar professional training course.
15. There have been suggestions that lawyers need to learn computer programming in order to safeguard and enhance their careers, but we do not consider that to be essential. Far more important is an understanding of the principles that underpin modern technologies, an ability to apply those technologies in the workplace and an ability to contribute to the discussion on the merits and ethics of the use of these technologies.
16. We predict the emergence of new types of roles within the sector, including roles that straddle technology and law, as it adjusts to offer broader services powered by technology. LexisNexis UK, for example,

⁸¹³ An example of such a task is due diligence undertaken as part of the acquisition of a company.

employs data scientists (specialising in extracting value and insight from legal data), knowledge engineers (legally trained professionals specialising in creating programs built around legal logic) and software developers. These new roles help us build products that provide predictive outcomes based on legal data such as Lex Machina and Ravel Law, as well as algorithm-driven products such as LexisDraft⁸¹⁴ that help lawyers optimise drafting material.

Question 8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

17. Some key implications of the uses of artificial intelligence described in paragraph 5 above (and which may also have relevance to others sectors beyond legal) include, but are not limited to, the following:

- a. Bias and discrimination in automated decision taking; and
- b. Unregulated self-service pathways.

Bias and discrimination in automated decision taking

18. The use of computers to make decisions on issues affecting the lives of human beings is not new but will become more common as artificial intelligence improves. The technology has the potential to introduce efficiencies into manual processes ranging from insurance claims to court sentencing⁸¹⁵.

19. The primary concern around the use of the technology in this way is that algorithms may contain biases or inherent assumptions that can be hard to detect and which may lead to discrimination. Such biases may originate in the data used to train the system, in data that the system processes during its period of operation, or in the person or organisation that created it. There are additional risks that the system may produce unexpected results when based on inaccurate or incomplete data, or due to any errors in the algorithm itself.

20. Legislation addressing automated decisions already exists. The Data Protection Act 1998 creates a right for individuals to prevent a decision being taken automatically. It also requires organisations to inform the individuals where an automatic decision has been taken and give them the right to request a review of the decision. These provisions will be replaced with expanded, but largely similar, provisions in May 2018 when the General Data Protection Regulation comes into effect.

Unregulated self-service pathways

⁸¹⁴ <http://www.lexisnexis.co.uk/en-uk/products/lexisdraft.page>

⁸¹⁵ See, for example, Sent to Prison by a Software Program's Secret Algorithms, New York Times, 1 May 2017.

21. A self-service pathway is a tool, usually web-based, that provides solutions through the use of chatbot-type technology.
22. They already exist, and are expected to become more prevalent, in a range of knowledge-intensive fields including in the legal industry for the provision of automated legal advice. An example of this is the DoNotPay bot referred to in paragraph 3 above.
23. The benefit of such tools in the legal sector is wider consumer access to legal solutions. In the absence of the tool, many would not seek professional legal advice and may be denied access to the legal system.
24. However, the easier they become to create and implement, the more likely it is that some self-service solutions may offer poor or incorrect advice, to the detriment of the user.
25. The appropriate way to deal with this issue is through sectoral regulatory bodies. Such bodies should be considering these issues now with a view to assessing how to harness the benefits that self-service pathways can offer while minimising the risks and increasing public confidence.

Question 9: In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

26. It should be recognised from the outset that some degree of non-transparency may be unavoidable. As explained in paragraph 1 above, artificial intelligence can potentially complete complex data-processing and analytical tasks that are beyond human capabilities. It is also able to learn and to modify its behaviour. As a result of this complexity, there may be a natural limit beyond which it is not possible to comprehensively deconstruct or analyse the intricate workings of algorithms in order to understand why they have reached the conclusions they have or behaved in a certain way.
27. Irrespective of this limitation, in considering these issues an appropriate balance must be struck between the protection of commercially sensitive information and the protection of consumers' fundamental rights.
28. At the same time, there must be recognition that artificial intelligence comprises a vast spectrum of applications. Some of these will be innocuous and low risk (for example, a chatbot that helps users to identify recipes⁸¹⁶). At the other end of the scale, some may be used for purposes that have serious long term implications⁸¹⁷.

⁸¹⁶ See, for example, <http://www.foodnetwork.com/site/apps/chatbot> (accessed 6th September 2017).

⁸¹⁷ See, for example, footnote 11 above.

29. Any proposed regulation in this area must therefore be risk-based. Transparency and accountability will be necessary where artificial intelligence makes decisions, or is used as part of decision-making processes, that have a material impact on the lives of individuals or where the risks of errors may cause serious harm.
30. Examples of measures that might be considered include requirements that:
 - a. certain organisations produce an assessment of the potential consequences of the application of their technology and take appropriate actions regarding any implementation of their technology based on this assessment; and/or
 - b. algorithms that are identified as being higher-risk, either due to their nature or intended use, must be registered with an independent body.
31. The independent body would need to be staffed with adequately qualified and experienced personnel and be empowered to make appropriate assessments of algorithms, and to take enforcement action, where required.

Question 10: What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

32. The Government's approach to artificial intelligence should comprise two strategic objectives.
33. Firstly, research and development into artificial intelligence should be supported to ensure that the United Kingdom can play a leading role in innovation. The Government has already committed to related technologies such as connected and autonomous vehicles. This commitment should be extended to other autonomous technologies. As part of this, the Government should ensure that digital education is prioritised and enhanced at all levels, including professional training.
34. Secondly, more research is needed on the best way to safeguard society against the challenges and risks that artificial intelligence presents. The approach should not be one-size-fits-all and should be based on risk: artificial intelligence applications that help a consumer answer a simple question online cannot be treated in the same way as a system that automatically rejects a benefit application. Equally, neither of these examples can be treated in the same way as, for example, lethal autonomous weapons.

35. Therefore, regulation that addresses the outcomes of artificial intelligence⁸¹⁸, or its use in specific sectors, is likely to be more effective than any attempt to regulate artificial intelligence *per se*.
 36. Broad regulation in response to general concerns about the impact of artificial intelligence should be avoided. The Government should instead employ a risk-based approach to regulation. The first step should be to identify foreseeable implications of artificial intelligence that pose the greatest risks and respond with appropriate proposals for regulation.
 37. We endorse the recommendations of the Science and Technology Committee⁸¹⁹ that a robotics and autonomous systems (RAS) leadership council should be formed together with a standing commission on artificial intelligence. In addition to making recommendations for appropriate regulatory reform, these bodies should focus on developing principles and guidelines to shape the short to medium-term development of artificial intelligence.
 38. The principles should seek to enshrine values, among other things, of openness, humanity, transparency, fairness, privacy, and security in the research and development of artificial intelligence.
- We would welcome the opportunity to provide further evidence, either in written or oral form, if it would assist the Committee.

LexisNexis UK

6 September 2017

⁸¹⁸ See, for example, the provisions of the Data Protection Act 1998 (and, when in force, the General Data Protection Regulation) on automated decision taking referred to in paragraph 20.

⁸¹⁹ Fifth Report of Session 2016-2017, Robotics and artificial intelligence (accessed 6th September 2017).

Liberty – Written evidence (AIC0181)

About Liberty

Liberty (The National Council for Civil Liberties) is one of the UK's leading civil liberties and human rights organisations. Liberty works to promote human rights and protect civil liberties through a combination of test case litigation, lobbying, campaigning and research.

Liberty provides policy responses to Government consultations on all issues which have implications for human rights and civil liberties. We also submit evidence to Select Committees, Inquiries and other policy fora, and undertake independent, funded research.

Liberty's policy papers are available at
<http://www.liberty-human-rights.org.uk/policy/>

Contact

Corey Stoughton
Advocacy Director
Laura Hickie
Campaigns Co-ordinator
Sam Hawke
Advocacy and Policy Officer

Rachel Robinson
Advocacy Manager
Silkie Carlo
Senior Advocacy Officer
George Wilson
EU Law and Policy Specialist

Gracie Mae Bradley
Advocacy and Policy Officer

Introduction

1. Liberty welcomes the establishment of the Select Committee on Artificial Intelligence and the opportunity to submit evidence to its inquiry.
2. The technological revolution is transforming society in numerable ways. In light of current and near-future seismic technological shifts, Liberty has expanded the scope of its work to include a new programme on Technology and Human Rights. We seek to provide the Committee with a brief human rights analysis of some of the great challenges and opportunities that artificial intelligence presents to the UK.
3. In this submission, Liberty wishes to focus on the following key issues:
 - a. AI, privacy and data protection
 - b. AI and equality rights
 - c. AI, transparency and accountability
 - d. AI and weaponry

The pace of change

4. The promise of scientific progress, clinical rationality and increased efficiency means AI is regarded as a highly desirable technology within

relevant sectors. There is a clear rush to deploy AI in various areas of research and public life, making the pace of change rapid.

5. The rapid pace of change perhaps explains some of the most controversial current uses of AI in the UK. AI is already being used for predictive policing by Kent Police, for health research on NHS patient data by Google DeepMind, and for online advertising including political advertising.
6. Such a rapid pace of change often means that fundamental considerations such as the impact on human rights, data protection, transparency and public consent are neglected. This risks a silent deterioration of core values in the name of scientific 'progress'.

Techno-optimists and techno-pessimists

7. Whilst a tangible dichotomy in practice, the distinction between 'techno-optimists' and 'techno-pessimists' is not a helpful one. Both terms suggest a somewhat deterministic role of technology in society, with stakeholders predicting rather than determining what technologies are used, when they are used, and the outcomes.
8. In Liberty's view, it is paramount to actively uphold the rule of law and the human rights framework provided by the Human Rights Act 1998 throughout the ongoing technological revolution. We firmly believe that human rights laws can play a leading role in safeguarding rights and liberties during this period of great change.
9. Human rights laws should not only guide the passage of new technologies into society – they should be hardwired into those technologies. Software should be designed with privacy, freedom of expression, accountability and civil liberties in mind as normative principles – ensuring fundamental rights underpin the digital sphere. 'Rights by design' may be more challenging where AI is concerned, as the software organically learns and adapts in response to its environmental input.

Summary

- **Privacy:** There is no tension between privacy or indeed any fundamental rights and socially beneficial scientific progression. Privacy is not a necessary sacrifice for positive applications of AI.
- **No discrimination:** Whilst data science clearly has the potential to illuminate and ameliorate discrimination, it is unclear why AI per se would be necessary for such progression. AI systems should not be reified as objective or decontextualised from the social context and ideologies within which they are constructed. Training datasets, and any social dataset used to train AI programmes, must be carefully assessed and controlled for patterns of historical and ongoing bias, or sampling deficiencies, to avoid the perpetuation or creation of discrimination and inequalities.
- **Diversity:** It is vital for the success of AI that workforces in the tech sector are representative of the diversity of experience and backgrounds within the society they seek to operate in.

- **Democracy:** The public should be informed of the areas in which AI is being applied and how it is being applied, whether in public policy or in relation to individual decisions. Liberty supports a growing public debate on the topic of AI.
- **Transparency:** AI systems should be transparent, open-source where possible, their functioning intelligible, their operation subject to democratic oversight, and both the systems' and their developers' decision-making accountable. AI-related decisions that engage the rights and liberties of individuals should always be challengeable.
- **Accountability:** AI and automatic processing must not be the sole basis for a decision which produces legal effects or engages the rights of any individual.
- **Stop Killer Robots:** Liberty joins the call for a ban, by way of international treaty, on lethal autonomous weapons systems that lack meaningful human control.

AI, privacy and data protection

10. Many AI systems are built on 'big data', whether 'open data' (publicly available data) or personal data. Some AI systems are built on smaller training datasets. Data protection, privacy and related rights must be closely regarded alongside the development of AI. In Liberty's view, it would be beneficial to incorporate these topics into computer science and related academic programmes.
11. Personal data is often the fuel for AI – whether for research, commercial products, or personalised services. The unlawful data sharing of 1.6 million identifiable patient records by the Royal Free London NHS Trust to a Google AI start-up, Google DeepMind, is a prime example of the risks to basic rights in the rush for AI. There is little personal information that is more profoundly private than medical information. Liberty is deeply concerned by the effect that this reprehensible data sharing has had on patients. We are providing free legal advice to a number of Royal Free patients who have contacted us seeking help, having lost confidence in the confidentiality they are entitled to in the course of their healthcare.
12. Seeking to build AI tools by training software with personal data received in breach of the law is needlessly reckless. Privacy and consent are not only pillars of our democracy, the rule of law, public health, but they are also essential for technological innovation. There need be no conflict between privacy and innovation – innovators must simply respect the rule of law and human rights in the course of advanced software development.
13. The shrinking of the private sphere and the growth of a surveillance society are broad concerns amplified by the increasing AI applications that are fuelled by personal data – even when data is lawfully exchanged. For

example, the increasing use of virtual personal assistants and growing personalisation functions in services requires software to learn from people by pervasively ingesting their data, effectively surveilling them. The BBC's head of digital partnerships recently said:

*"Just by listening to the voices in the room, your TV could automatically detect when there are multiple people in the living room, and serve up a personalised mix of content relevant to all of you in the room."*⁸²⁰

This data exchange may be well-intended and deemed to be to the user's benefit. Accordingly, the data exchange could be constructed in a lawful way – for example, if the service were optional and on the basis of fully informed consent. Even so, it remains that the normalisation of pervasive monitoring, passive quantification and intelligent personalisation risks reshaping society in ways unconsidered. Never before have human societies been monitored and quantified in this way – pervasive monitoring leads to self-monitoring, whether conscious or unconscious, and inhibits the development of personalities and ideas. Truly private spaces risk being eradicated, even from the home. Constant personalisation creates risks too, both for a free press and freedom of thought. The benefits of exposure to diverse media sources risk being limited by the echo chambers artificially constructed around each person, fuelling radicalism and social divides.

14. This risk is compounded by the pervasive suspicionless surveillance people are subjected to by the State, as individuals can never be sure that the data they generate is *only* being used for personalised services, etc., and not also being aggregated by the authorities. Already, our phone call records, text message records, GPS location data, Automatic Number Plate Recognition (ANPR) data, TfL data, travel data, banking data, and internet browsing records – not to mention unnamed 'bulk personal datasets' – are hoarded and processed. The State's secretive approach to web logging and data gathering may cause people *"to feel that their private lives are the subject of constant surveillance"*⁸²¹ – even when services promise not to pass data on to a third party, State surveillance is always exempt and its shadow hangs over society, chilling free expression.

15. In addition, unique rights and ethics issues may still arise where data is interpreted and acted on using AI – again, even where data is exchanged

⁸²⁰ Future BBC iPlayer could tell who is in the room and notice when the children have gone to bed – Anita Singh, The Telegraph, 19 Aug 2017: <http://www.telegraph.co.uk/news/2017/08/19/future-bbc-iplayer-could-track-familys-movements/> (accessed 21 Aug 2017).

⁸²¹ Joined Cases C-203/15 and C-698/15, *Tele2 Sverige AB v Post- och telestyrelsen, and Secretary of State for the Home Department v Tom Watson*, 21 December 2016, para. 100

lawfully. For example, Facebook has started using AI to identify users deemed at risk of suicide, in order to launch interventions.⁸²² The technology that exists to do this is increasingly advanced, and now predictive.⁸²³ There is no evidence that suggests this type of 'intelligent' monitoring is in the best interests of vulnerable peoples' mental health and we are concerned that a shrinking private sphere may indeed deter people from seeking social support and a safe space to freely express themselves.

16. This 'intelligent' processing of varied information should be subject to the data protection principle of fair and lawful processing. Liberty welcomes the clarity that will be provided by the EU General Data Protection Regulation (GDPR), which we anticipate will be passed into UK law via the forthcoming Data Protection Bill 2017, on the matter of specific, informed and unambiguous consent for data processing⁸²⁴ and the right not to be subjected to automated profiling.⁸²⁵

AI and equality rights

17. AI systems are built from programmed rules, training datasets, and learned rules from past outcomes. Inherent in the generation of AI systems, therefore, is potential for in-built bias either from the developers, programmed rules, the training datasets, or past outcomes.
18. System developers may programme rules in AI systems that are influenced by unconscious biases. Constructing AI systems requires the selection of certain features and whilst this relies on the mathematical expertise of developers, the process will also inevitably be influenced by their intuition, underlying worldview, culture, and motivations.⁸²⁶ Thus, transparency of the function of AI systems and close monitoring of how they operate in the real world is very important.
19. Training data, or in fact any data collected from society, may be reflective of patterns of discrimination and existing inequalities: "*to the extent that society contains inequality, exclusion or other traces of discrimination, so too will the data*".⁸²⁷ Social data come with complex histories, which may silently haunt the logic underpinning social policy if uncritically used.

⁸²² *Facebook Using Artificial Intelligence to Help Suicidal Users* – Aatif Sulleyman, The Independent, 2 March 2017.

⁸²³ *Artificial Intelligence Can Now Accurately Predict Suicide Attempts Two Years in Advance* – Paul Tamburro, Crave, 3 March 2017

⁸²⁴ GDPR Article 4(11) and Recital 32

⁸²⁵ GDPR, Article 4(4); GDPR, Article 13 (2)(f), and in particular, GDPR, Article 22(1)

⁸²⁶ *Algorithmic paranoia and the convivial alternative* – Dan McQuillan, Big Data & Society, July-December 2016, p.4

⁸²⁷ *European Union regulations on algorithmic decision-making and a "right to explanation"* – B. Goodman, S. Flaxman, Aug 2016, p.3

Patterns of social inequalities can be perpetuated through algorithmic processes – for example, Google advertises highly paid jobs to men more often than to women.⁸²⁸ The institutionalisation of AI systems may have more serious effects still.

20. Algorithmic profiling involves categorising and analysing people on the basis of a variety of categorisations and group memberships.⁸²⁹ This is not only an issue when sensitive data, such as race, gender, health, religion, etc., are used for profiling. Even if race is prohibited as a category of profiling data (as may be the case under the GDPR), combinations of other categories of data can unintentionally serve as a proxy for race. For example, *“if a certain geographic region has a high number of low income or minority residents, an algorithm that employs geographic data to determine loan eligibility is likely to produce results that are, in effect, informed by race and income.”*⁸³⁰ This may be important to consider in uses of AI by law enforcement. For example, Kent Police use predictive policing software, PredPol, with little transparency – it is not publicly known what categories of data are processed. Our engagement with police forces about the potential for discriminatory biases in various algorithms they use has thus far revealed a concerning disregard for the issue.⁸³¹
21. In addition, a predictive policing tool is likely to be based on large amounts of existing policing and crime data – but if this data reflects socio-economic, geographic or racially based discriminatory policing, those biases risk being entrenched in the tool the data seeks to produce, as early evidence suggests.⁸³²
22. Durham Police is using AI software, the Harm Assessment Risk Tool (‘HART’) in bail decisions, with little transparency. There is evidence that data science has perpetuated discrimination in criminal justice in the US. A recidivism algorithm called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions, by Northpointe, Inc.) was found to be twice as likely to incorrectly judge black defendants as at high risk of reoffending than white defendants.⁸³³ This is despite race not being one of the categories of information ingested. Without open source algorithmic transparency, we cannot know exactly how or why the algorithm came to these conclusions. Such decision making should be challengeable, subject

⁸²⁸ *Artificial Intelligence’s White Guy Problem* – Kate Crawford, The New York Times, 25th June 2016

⁸²⁹ *European Union regulations on algorithmic decision-making and a “right to explanation”* – B. Goodman, S. Flaxman, Aug 2016, p.3

⁸³⁰ *European Union regulations on algorithmic decision-making and a “right to explanation”* – B. Goodman, S. Flaxman, Aug 2016, p.4

⁸³¹ *Misidentification and improvised rules - we lift the lid on the Met's Notting Hill facial recognition operation* – Silkie Carlo, Liberty, Aug 2017

⁸³² *“Stuck in a Pattern: Early evidence on ‘predictive policing’ and civil rights,”* - David Robinson and Logan Koepke, Upturn, August 2016

⁸³³ *Machine Bias* - Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica, May 23, 2016

to an adversarial court proceeding - this opaque AI application is discriminatory and highly inappropriate.

23. Liberty has made requests for information to Durham Constabulary about HART under the Freedom of Information Act, which have been rejected – in our view, wrongly – under Section 21 (information reasonably accessible to the applicant by other means). We have recently resubmitted those requests. We were instead directed to an academic paper and Durham Constabulary’s written submission to the Science and Technology Committee’s inquiry into algorithms in decision making. Regarding the risk of discriminatory patterns in data being silently reproduced in algorithms, Durham Constabulary wrote in that submission,

“It would be wrong, and the error rates would increase, if the model failed to reflect reality. Before concluding that algorithms should therefore be viewed as biased, it is necessary to consider whether human judgement is more or less biased.”⁸³⁴

This attitude reflects a disregard of the duty police have to ensure equality and fairness in policing. ‘Reflecting reality’ is not where the bar for success should be set. The threshold for adopting AI in criminal justice and law enforcement should not simply be whether one function of the software exceeds human functioning on some measure – software must be considered holistically, accounting for new risks and long-term impact. Data analysts should, at the very least, attempt to discover and control for biases in existing datasets before using them to train AI tools for live deployment in the criminal justice system where they risk being embedded, reified and obscured from accountability. Seemingly progressive AI applications may produce socially regressive output and must not automatically be attributed as objective, or divorced from ideology.

24. Another reason that algorithms may not produce fair decisions for minority groups is that there may be too small a sample of data from which to generate predictions with confidence. For this reason, minorities are sometimes oversampled in public policy research.⁸³⁵ An example of inadequate training data resulting in discrimination was seen in Google’s photo app, which classified black people as gorillas. Similar examples of algorithmic discrimination include Nikon software reading photos of Asian people as blinking and HP webcam software having difficulty recognising

⁸³⁴ Written evidence submitted by Durham Constabulary (ALG0041; para. 22) in response to the Science and Technology Committee’s inquiry into algorithms in decision making – April 2017

⁸³⁵ *European Union regulations on algorithmic decision-making and a “right to explanation”* – B. Goodman, S. Flaxman, Aug 2016, p.4

users with dark skin tones.⁸³⁶ Low representation in training data could be a possible explanation for higher inaccuracy rates for the identification of women by facial recognition software.⁸³⁷ Oversampling may be a solution in some circumstances, but if doing so requires increased surveillance or data collection on a particular group it could raise serious legal and ethical issues. We caution against subjecting a portion of society to increased surveillance for a technocratic 'greater good' – particularly for groups that are typically marginalised.

25. The performance of AI systems must be carefully tested before systems are operational, and continually monitored, as discriminatory flaws may not be easily discoverable and too often only come to attention once they already are having negative effect. Disturbingly, Liberty has witnessed a wilful ignorance from police towards potential discriminatory flaws, who seem confident in the professed objectivity of such software. For example, the Metropolitan Police told us they had no intention to monitor demographic accuracy (or even to ask the vendor if it had been tested for) in their use of facial recognition software. In addition, the Commissioner of the Metropolitan Police failed to respond to, or even acknowledge, a letter from Liberty and 12 other rights and race equality groups raising concerns about accuracy bias.⁸³⁸
26. Whilst the risks of embedding patterns of discrimination in AI are clear, it is possible to use data and design to control for and thus reduce discrimination. There is a view in the US that, if care is taken to minimise the possibility of biases or inaccuracies in the data, AI "*has the potential to improve aspects of the criminal justice system, including crime reporting, policing, bail, sentencing, and parole decisions*".⁸³⁹ Whilst data science clearly has the potential to illuminate and ameliorate discrimination, it is unclear why AI per se would be necessary for such progression. Clearly, the COMPAS case study was not successful.
27. Since software, including AI, reflects the values of its creators it is important that workforces in this sector are representative of society. It is particularly urgent address the under-representation of women - only 17% of those working in technology in the UK are female and just 7% of students taking computer science A-level courses are female.⁸⁴⁰

⁸³⁶ [Artificial Intelligence's White Guy Problem](#) – Kate Crawford, The New York Times, 25th June 2016

⁸³⁷ [Face Recognition Performance: Role of Demographic Information](#) - [Brendan F. Klare](#) ; [Mark J. Burge](#) ; [Joshua C. Klontz](#) ; [Richard W. Vorder Bruegge](#) ; [Anil K. Jain](#), [IEEE Transactions on Information Forensics and Security](#) (Volume: 7, Issue: 6, Dec. 2012)

⁸³⁸ [Misidentification and improvised rules - we lift the lid on the Met's Notting Hill facial recognition operation](#) – Silkie Carlo, Liberty, Aug 2017

⁸³⁹ [Preparing for the Future of Artificial Intelligence](#) – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.14

⁸⁴⁰ Women In Tech (womenintech.co.uk), accessed Sept 2017

AI, transparency and accountability

28. It is important that decision making is accountable in a democracy – both of AI creators and their creations. However, accountability for their creations can be frustrated by several factors: the highly complex, multi-dimensional nature of many processing systems that are not easily interpretable if at all; the prevalence of commercial, proprietary systems and of secret systems (e.g. in state intelligence); the inability of probabilities to offer explanations for output beyond the assumption that past associations between data will be replicated in the future; and the inaccessibility of source code and complex algorithmic rules to many individuals, even where it is published.
29. The internal machinations of AI systems are often highly complex, opaque and sometimes near-impossible to 'reverse-engineer'. Although AI processing of big data can reveal previously invisible relationships, the processing itself is inherently opaque. AI systems are rarely designed to provide explanations for the output they produce or decisions they make. AI-based social decision-making "*renders individuals unable to observe, understand, participate in, or respond to information gathered or assumptions made about them*" and has been described as "*antithetical to privacy and due process values*".⁸⁴¹ It can even be argued that AI in social decision making can be "*authoritarian (...) in that it eludes democratic oversight and, so far, evades a social discourse capable of challenging its teleology*".⁸⁴² We must be alive to the potential that authoritatively 'objective', opaque calculations behind social policy in the future may evade challenges based on human reasoning, disempower the public and shift the relationship between the citizen and state.
30. Furthermore, AI systems are increasingly designed to learn, adapt and improve so their processes may change during deployment. Such systems pose "*perhaps the biggest challenge – what hope is there of explaining the weights learned in a multilayer neural net with a complex architecture?*"⁸⁴³ This opacity, combined with perceptions of AI as producing perfect, rational, if unknowable calculations, creates a situation in which important decisions may be readily accepted but too reluctantly questioned or scrutinised. It is certainly true that "*algorithmic vision derives authority from its association with science (...) an aura of neutrality and objectivity,*

⁸⁴¹ *Prediction, pre-emption, presumption: How big data threatens big picture privacy* – J. Earle & I. Kerr, Stanford Law Review Online 66:65, 2013

⁸⁴² *Algorithmic paranoia and the convivial alternative* – Dan McQuillan, Big Data & Society, July-December 2016, p.5

⁸⁴³ *European Union regulations on algorithmic decision-making and a "right to explanation"* – B. Goodman, S. Flaxman, Aug 2016, p.7

which can be used to defend against the critique that they carry any social prejudice.”⁸⁴⁴

31. An additional risk of opaque AI systems is that they may have unintended consequences or malfunction, without being readily noticed or understood. This idea is sometimes the source of dystopian science fiction, but could have some more everyday negative impacts. The recently agreed Alisomar Principles on AI state that developers must provide transparency to understand failures, plan for/mitigate catastrophic risks, and ask questions such as, *“how can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?”⁸⁴⁵*
32. Researchers should design open-source AI systems where possible, and always maintain the ability to provide transparency as to their output and explanations for decisions made. Human interpretability is especially important where AI is used in public decision making and applications that could have human rights engagements or require ethical consideration. Both AI systems and their creators must remain accountable.
33. It is important to maintain the same standards of accountability and transparency when using AI – we believe AI, particularly where it interacts with individual’s rights, should be rejected where transparency is not possible. It would be unacceptable to consider AI systems as immune from accountability.
34. It is important that the public is informed in which areas AI systems are being used, that changing behaviours of adaptive AI systems are closely monitored, and that AI decision-making is always transparent. The Science and Technology Committee described decision-making transparency as one of the *“key ethical issues requiring serious consideration (...) [and] ongoing monitoring”*.⁸⁴⁶ Similarly, a US Executive Committee on Technology recommended: *“As the technology of AI continues to develop, practitioners must ensure that AI-enabled systems are governable; that they are open, transparent, and understandable”*.⁸⁴⁷
35. Where it is argued that AI systems or their controllers cannot be transparent or provide explanation for decisions, for example where they are used in the intelligence community, we call for the maximum possible public transparency with full transparency allowed in a closed independent review or adversarial procedure. This is important to verify that the

⁸⁴⁴ *Algorithmic paranoia and the convivial alternative* – Dan McQuillan, Big Data & Society, July-December 2016, p.4

⁸⁴⁵ See Appendix I

⁸⁴⁶ *Robotics and artificial intelligence* – Science and Technology Committee, Sept 2016, p.36

⁸⁴⁷ *Preparing for the Future of Artificial Intelligence* – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.4

subject's rights and freedoms are safeguarded, particularly where a system has legal or other significant effects on a subject.

GDPR

36. In limited circumstances, EU citizens may soon have the right not to be subject to algorithmic decisions that would significantly or legally affect them. Article 22 of the EU's new General Data Protection Regulation, set to take effect from 2018, states:

*"The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."*⁸⁴⁸

This principle does not apply if the decision is authorised by EU or state law so long as the data subject's rights, freedoms and legitimate interests are safeguarded.⁸⁴⁹ It is also disapplied if it is necessary under a contract between the data controller and the subject,⁸⁵⁰ or the subject has given explicit consent.⁸⁵¹ Nevertheless, this Article may prohibit private corporations using AI in various applications.

37. This Article leaves open the possibility for a citizen to be subject to a decision based *solely* on automated processing, that produces legal effects, without the right to human intervention or to express their view and contest the decision, so long as the state or EU has safeguards in place to protect the subject's rights, freedoms and legitimate interests. It is hard to envisage what safeguards could effectively protect against the risks inherent in this model.

38. In Liberty's view, individuals should have the right not to be subject to a decision by the state that is based solely on automatic processing and which produces legal or other significant effects for them.

39. Article 22 of the GDPR further offers EU citizens 'the right to explanation' regarding an algorithmic decision made about them with consent or under contracts. The right to explanation does not apply to the State's use of AI. Article 22(3) states that the subject maintains,

⁸⁴⁸ *General Data Protection Regulation, Article 22(1)*

⁸⁴⁹ *General Data Protection Regulation, Article 22(2)(b)*

⁸⁵⁰ *General Data Protection Regulation, Article 22(2)(a)*

⁸⁵¹ *General Data Protection Regulation, Article 22(2)(c)*

“(...) the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.”⁸⁵²

Furthermore, Articles 13 and 14 state that subjects have the right to be given *“meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing”*.⁸⁵³ Article 12 states that communication with data subjects must be in *“concise, transparent, intelligible and easily accessible form”*.⁸⁵⁴ These regulations go some way to addressing concerns about transparency around proprietary AI systems, as well as concerns about the accessibility of transparency mechanisms such as open source code – but, given the exceptions, do not fully satisfy our concerns about the transparency of state AI systems. Transparency is arguably even more important where the decision originates from the State and has significant or legal effects for the subject.

AI & lethal autonomous weapons

40. Lethal autonomous weapons systems (LAWS) could potentially identify and kill a target without human intervention. The Government has said it will not develop LAWS, but defines LAWS with a very vague and futuristic view of ‘autonomy’: *“An autonomous system is capable of understanding higher level intent and direction”*.⁸⁵⁵ Furthermore, the Government has opposed proposals for an international ban on the development of LAWS⁸⁵⁶, or the development of any guidelines or additional legislation.⁸⁵⁷ Both the UK and US cite international humanitarian law and ongoing international discussion as key to regulating the use of LAWS.^{858, 859}

41. Precursors to fully autonomous weapons are being developed and deployed by the UK as well as the US, Russia, China, Israel and South Korea. A drone called Taranis is being developed by the UK’s MoD and BAE

⁸⁵² *General Data Protection Regulation, Article 22(3)*

⁸⁵³ *General Data Protection Regulation, Article 13(2)(f) and Article 14(2)(g)*

⁸⁵⁴ *General Data Protection Regulation, Article 12*

⁸⁵⁵ *Joint Doctrine Note 2/11: The UK Approach to Unmanned Aircraft Systems*, Ministry of Defence, 30 March 2011, para 205, p.14

⁸⁵⁶ *UK opposes international ban on developing ‘killer robots’* – Owen Bowcott, The Guardian, 13 April 2015, <https://www.theguardian.com/politics/2015/apr/13/uk-opposes-international-ban-on-developing-killer-robots>

⁸⁵⁷ *Statement on Lethal Autonomous Weapons Systems to the CCW Meeting of the High Contracting Parties*, United Kingdom of Great Britain and Northern Ireland, 12–13th November 2015

⁸⁵⁸ *Preparing for the Future of Artificial Intelligence* – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.3

⁸⁵⁹ *Statement on Lethal Autonomous Weapons Systems to the CCW Meeting of the High Contracting Parties*, United Kingdom of Great Britain and Northern Ireland, 12–13th November 2015

Systems, and has reportedly been tested to autonomously locate and engage targets.^{860 861}.

42. The US has more openly embraced autonomy in weapons systems. A US Executive Committee on Technology stated: "*The United States has incorporated autonomy in certain weapon systems for decades, allowing for greater precision in the use of weapons and safer, more humane military operations (...) Nonetheless, moving away from direct human control of weapon systems involves some risks and can raise legal and ethical questions.*"⁸⁶²
43. UK Government has committed to keeping weapons under human control, but has not specified what it understands by human control. For example, 'human control' may mean authorising weapons release when prompted by a system to do so. A more appropriate goal may be 'meaningful human control' – a term that requires development but is orientated on the premise that humans must exert meaningful, cognitive control of combative actions and perform important decisive functions.
44. Even retaining meaningful human control, the site of accountability for such advanced weapons use may be problematic. Military personnel may not fully understand the internal machinations of these highly advanced systems – particularly of such closely guarded national security technologies. Should they be held accountable for using equipment that, whilst advertised as more safe, precise and rational than human intervention, they do not entirely understand?⁸⁶³ Who would be held accountable, not just in law but on the political stage, for the malfunctioning of a LAWS? Clearly, maintaining transparency and meaningful human control of weapons systems' functioning will be vital to ensure proper accountability frameworks and the upholding of obligations under human rights law.
45. Liberty calls for a pre-emptive ban on the development and use of lethal autonomous weapons that do not involve meaningful human control. We concur with Human Rights Watch and industry leaders that such a pre-emptive ban is necessary now.^{864 865} We support the Government's

⁸⁶⁰ *Anglo-French UCAV Study Begins To Take Shape* – Toby Osborne, Aviation Week, 4th Feb 2016, <http://aviationweek.com/defense/anglo-french-ucav-study-begins-take-shape>

⁸⁶¹ *The United Kingdom and lethal autonomous weapons systems* – Article 36, April 2016, p.2

⁸⁶² *Preparing for the Future of Artificial Intelligence* – Executive Office of the President, National Science and Technology Council Committee on Technology, Oct 2016, p.37

⁸⁶³ See also evidence to the Science and Technology Committee by Richard Moyes, Article 36, cited in *Robotics and artificial intelligence* (Sept 2016), para/ 56, p.21

⁸⁶⁴ *Killer Robots* – Human Rights Watch, accessed 25.1.16. See:

<https://www.hrw.org/topic/arms/killer-robots>

⁸⁶⁵ *Written evidence submitted to the Science and Technology Committee's inquiry on Robotics and Artificial Intelligence* (ROB0062) – Google DeepMind, May 2016

Liberty – Written evidence (AIC0181)

commitment to maintaining human oversight and control over the use of force, but stress that such control must be *meaningful*.

6 September 2017

Chrissie Lightfoot, Michael Butterworth, Ms Joanna Goodman and Dr Paresh Kathrani and Dr Steven Cranfield – Written evidence (AIC0104)

Chrissie Lightfoot, Michael Butterworth, Ms Joanna Goodman and Dr Paresh Kathrani and Dr Steven Cranfield – Written evidence (AIC0104)

[Submission to be found under Ms Joanna Goodman](#)

Dr Conor Linehan and Dr Dan O'Hara, Professor Shaun Lawson and Dr Ben Kirman – Written evidence (AIC0127)

Dr Conor Linehan and Dr Dan O'Hara, Professor Shaun Lawson and Dr Ben Kirman – Written evidence (AIC0127)

[Submission to be found under Dr Dan O'Hara](#)

Professor Rosemary Luckin – Supplementary written evidence (AIC0246)

House of Lords AI Select Com Additional Evidence about the benefits of AI when applied to the education of SEND students

There are a range of ways in which artificial intelligence can be used to support the education of students with special educational needs. For example, the use of natural language processing to enable the development of voice activated interfaces can be helpful for students with physical disabilities that restrict their use of other input devices, such as keyboards.

The combination of artificial intelligence and other technologies such as virtual and augmented reality can help students with physical and learning disabilities to engage with virtual environments and take part in activities that would be impossible for them in the real-world.

Virtual reality becomes 'intelligent' when it is augmented with AI technology. AI might be used simply to enhance the virtual world, giving it the ability to interact with and respond to the user's actions in ways that feel more natural. Or, drawing on Intelligent Tutoring Systems, AI might also be integrated to provide on-going intelligent support and guidance to ensure that the learner engages properly with the intended learning objectives without becoming confused or overwhelmed. Virtual pedagogical agents might also be included, acting as teachers, learning facilitators, or student peers in collaborative learning 'quests'. These agents might provide alternative perspectives, ask questions, and give individualised feedback. In addition, intelligent synthetic characters in virtual worlds can play roles in settings that are too dangerous or unpleasant for learners. For example, FearNot is a school-based intelligent virtual environment that presents bullying incidents in the form of a virtual drama. Learners, who have been victims of bullying, play the role of an invisible friend to a character in the drama who is bullied. [The learner offers the character advice about how to behave between episodes in the drama and, in so doing, explores bullying issues and effective coping strategies.](#)

AI can also help EdTech applications be more flexible, through, for example, deployment online, meaning that they can be available on personal and portable devices within, and beyond, formal educational settings.

The way that AI enables technology to be personalised to the individual needs of a learner can also make it beneficial for learners with special educational needs. Systems that use AI in this way include the sort of software produced by: Alelo in the US, who have been developing culture and language learning products since 2005 and specialise in experiential digital learning driven by virtual role play simulations powered by AI. Carnegie Learning produce the software that can support students with their

mathematics and Spanish studies. In order to provide individually tailored support for each learner the software must continually assess each student's progress. The assessment process is underpinned by an AI-enabled computer model of the mental processes that produce successful and near-successful student performance.

UK-based Century Tech, has developed a learning platform with input from neuroscientists that track students' interactions, from every mouse movement and each keystroke. Century's AI looks for patterns and correlations in the data from the student, their year group, and their school to offer a personalised learning journey for the student. It also provides teachers with a dashboard, giving them a real-time snapshot of the learning status of every child in their class.

Specific Educational Needs International Examples

1. The group called DoIT at the University of Washington (State of Washington) has been researching how to make every document on the internet available to people with Special Needs. They have also begun to use AI. [Click here](#).
2. AI is being used with students who have ADHD disability in work being done at Athabasca University. The long-term goal of this work is to develop an AIED (learning analytics) system that a) detects ADHD earlier than current models, b) improves the quality of diagnosis of ADHD c) educates instructors about methods that are effective for teaching students afflicted with ADHD, d) formatively and observationally measures competency improvements and challenges of ADHD students) engages/encourages ADHD students to study in an environment filled with anthropomorphic pedagogical agents. See: <https://journals.colostate.edu/analytics/article/view/131>.
3. [Guiding Technologies: A Temple University spin-off](#) based on NSF funded research is conducting intensive trials to use AI-enabled software to overcome problems in delivering Applied Behavior Analysis (ABA), the gold standard in treating developmental delays due to autism spectrum disorder (ASD) and intellectual challenges.
4. A range of work with people who have autism spectrum disorder. For example, using Pedagogical agents and personalized learning: https://link.springer.com/chapter/10.1007/978-3-540-27817-7_28
5. Systems that Leverage Big Data to help individual learners can also address special needs requirements. See for example, work with the [nStudy software system at Simon Fraser University](#)

9 January 2018

Dr Mike Lynch – Written Evidence (AIC0005)

The pace of technological change

- 1 What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Artificial Intelligence has now crossed the threshold for so called 'narrow AI' to solve a practical level of problems that was not possible before. This is as a result of the discovery in the late 80's of probabilistic self-learning systems which now display performance suitable enough to solve previously very difficult problems. This was made possible by algorithmic developments and the availability of very high computational power computers.

Narrow AI is likely to continue going forward as more data, computing power and more knowledge become available. However, it should be noted that success in narrow AI, defined as the ability to solve specific problems such as driving a car, or understanding speech, or replacing due diligence in a law firm, is very different to broad AI which is the ability to handle everything that comes to you from the real world.

A mistake that the general public often make is seeing broad AI as a continuum of narrow AI, where this is not the case. The great progress we are seeing at the moment in particular narrow AI applications does not mean that we are suddenly going to have sentient computers that can handle everything. Thus, a lot of the commentary in that area is rather misguided.

Factors that will affect things going forward include the technology itself and the ability to have large amounts of computation available. But the most important limitations will be the access to data to train the systems, as well as societal aspects like the possibility to have a legal liability framework and insurance. This is vital to allow these systems to actually be used. If insurance and legal liability aren't sorted out this will be a great hindrance to the technology being adopted.

- 2 Is the current level of excitement which surrounds artificial intelligence warranted?

The level of excitement in itself is warranted in that, for the first time, computers can do a series of tasks that previously could only be done by humans, often leading to an increase in the quality of those tasks because computers can see more, don't get bored as humans do and cost a lot less to operate. It is fair to

call this a revolution. Like many technology revolutions, the general population assume it is going to happen faster than it will and will probably underestimate the final impact of it. However, it should be noted that the phrase "AI" has become an overused marketing phrase where the majority of applications and companies talking about AI actually have very little AI ability and so there will be a lot of disappointments along the way.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

These systems will have a very significant impact on jobs. There have been various estimates, but we can probably assume that 30% of jobs will be affected. If you are an optimist, you think that there will be new roles created, but even in that scenario we also have to consider that the workforce will have to become much more adaptable, and their skills, particularly in areas such as problem solving will have to improve. It is likely that people will need to supplement their skills which will hark back to the fundamental pillars of high school education such as mathematics and English, without which they will find it very hard to adapt.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

It ought to be everyone will see some gains in that AI has the ability to affect so many areas, from personalized medicine, social care, transport policy, etc. It is a great enabling technology. The big difficulty of course is that unskilled or low skilled workers are going to be greatly under threat in terms of employment, and that is where it will be necessary for them to have some skills and be retrained.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Yes, it would be a very good idea for the public to be given a more realistic idea of AI, especially the difference between what really does exist, which is narrow

AI versus the broad AI of Science Fiction, and the fact that the two are not on a continuum.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

AI is a fundamental enabling technology and just as mechanization in the industrial revolution affected all areas of human economic activity, so it will be with AI. Therefore, it is really quite pointless to try to work out which sectors will be the most affected. It will also have a large effect on delivery of public services, so the question really misunderstands the fundamental nature of this technology. Any task where a human is involved in making a decision is potentially open to the effects of artificial intelligence.

Some people think that areas that require creativity or empathy may somehow be left out of the artificial intelligence revolution. This will probably be shown to be false, as it is quite likely that AI will be able to show empathy and a level of creativity.

There are many tasks that involve broad AI. For example, within the law, understanding how to navigate negotiating a deal will very much remain in the domain of humans although by contrast I expect the drudgery of going through hundreds of documents to check them could be done by a machine, as this is a narrow AI problem. This is a pattern we will see in most sectors where the broader, more creative work will stay suitable for humans but the more repetitive work is likely to be taken over by machines.

7. How can the data-based monopolies of some large corporations, and the 'winner- takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Data is everything in machine learning, which means whoever gets access to data can have a big advantage. As they gain a more consolidated position in the market, in turn they get access to more data, and so they can easily create a advanced competitively defensive position. Governments should be thinking about the concept of strategic data, which is data that has great value to it. For example, if the NHS were to give data to a machine learning company which then produces a very useful system with that data, the government needs to

realize that that has only been possible through the NHS data set. So, for example the company using that data might have to agree to have most favoured nation pricing for the NHS. At the moment, current government policy is sadly missing the understanding of strategic data and its importance to the UK economically but also to public services which will require the services produced by this data. Sadly, this debate is often eclipsed by the open data debate which is a rather academic debate that doesn't understand the economic effect of strategic data or indeed the fact that often the government and public bodies have strategic data which should not be used by commercial organisations to create systems which they will then be held to ransom to use.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

This is not my area of expertise

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

People who do not understand how the systems work often expect them to give explanations for their conclusions. The very power of these deep learning systems comes from the fact that they take together many, many subtle pieces of information to produce a very accurate conclusion. This inherently is not suited to giving simple explanations. In fact, simple explanation from any of these problems are an illusion. Even humans use the same subconscious weighing of very large amounts of information to come to a conclusion and often will give a reason which is a post-justification. However, various experiments show that these justifications are not the basis on which these decisions are made. It is very important to understand that explainability even in our current systems and society is often an illusion. What this means is that there is a tradeoff between accuracy of solving the problem (getting the right answer) and explainability. The most accurate answers will be given by the systems that are the least explainable. If one wished to impose explainability on a system then accuracy will fall. It is important to understand this most fundamental tenet because, for example, to impose a requirement of explainability on machine learning systems would greatly hold back the systems and their application and indeed the UK in this area.

However, it fair to say that in some areas for societal reasons we might want explainability – for example why a mortgage was turned down or why one

prisoner is released before another.

In these situations, it will be a question of deciding how much accuracy in the answer one would be prepared to give up in order to have some level of explainability. It is vital to understand that these two things are in balance and you cannot have one without the other. By the very nature of the problems that AI solves, 'intuition' is very much a part of getting the answer – in some cases, this is in opposition to explainability.

A big mistake that people make is to believe that explainability in deep learning systems is possible. At the moment, although it is a very interesting research area, it is not possible and therefore an ill-considered demand for explainability in all areas would actually hamper the UK greatly.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

The role of the government should be to sort out a framework for legal liabilities rather like the ones that allow airlines to function and therefore make sure that these systems are insurable. So, for example, driverless cars are likely going to have a lower error rate than human beings, therefore they should be easier to insure, but because human beings have precedence in their cases, it is likely to take some time for liabilities for driverless cars to come down to that level. That will greatly hold back the field so legislation in the area of risk or liabilities rather like that which allows other industries to function will be very important.

Artificial Intelligence systems will be hard to regulate without using Artificial Intelligence systems themselves, and there will be some areas where this has to happen sooner than others. If one looks at modern financial trading which is often now done by AIs, the only way to regulate them is by other AIs - humans do not have the ability to understand the complexity or the to do those calculations, so the regulatory area is very important.

One other aspect is strategically thinking about regulation. The US for example in its approach to copyright law worked in a way which greatly advanced the US internet sector. Therefore, regulation which allows that looks more forward more quickly to these changes for example in areas such as fintech and transport could aid the UK in becoming a leader in its field. This is often not about in any way lowering regulation but about being more forward looking. So, for example allowing a regulatory framework which allows driverless cars or more personalized medicine to be tested and used in the UK would greatly enhance the UK's position.

Learning from others

11. What lessons can be learnt from other countries or international organizations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

At the moment, the UK has been very forward looking in Artificial Intelligence. The Age of Algorithms report done by the Council for Science and Technology is one of the leading reports in area, which led to the formation of the Turing Institute, and actually institutions such as Cambridge University, the University College London and the Turing Institute are great sources of knowledge. It is hard to find other parts of the world that are actually more advanced in thinking in these areas than the UK.

27 July 2017

Dr Mike Lynch – Supplementary written evidence (AIC0230)

- 1. What, in your opinions, are the biggest opportunities and risks in the UK over the coming decade in relation to the development and use of AI?**

Opportunities

In terms of opportunities, AI is likely to be one of the most significant disruptions in the way public services are delivered and has very strong commercial opportunities. The UK, because of its very strong science base, is in a great position to flourish in this new world if it can turn its academic research into commercial impact and if public services can also adapt to new methods of working. In some cases, political bravery will be needed to overcome issues such as use of data in the health service where, although there are very significant benefits such as patient health, mortality and cost, unsubstantiated fears often lead to campaigns against data use and there will need to be a political will to overcome this.

Risks

As has been discussed on many previous occasions, there is likely to be significant disruption to the jobs market with a number of roles being disrupted or replaced by machines. There will be a move where some knowledge workers will be displaced by machines, although wisdom workers will still be becoming even more valuable. For those jobs displaced, there is reason to believe that new jobs will be created however this will lead to significant workplace skills and adaptability issues, and a likely period of transition which will be difficult.

One of the next key risks is the monopolisation of the AI market. The AI market relies on data to train the systems, so companies which get access to data more rapidly will take the lead and this will likely reinforce their position as they get more data, creating very strong monopolies. Those monopolies, if they were outside the UK, could be very detrimental to the UK economy. From that point of view, it is very important to make sure that the UK is very forward-looking in allowing access of data on a basis that is preferenced towards companies that will benefit the UK, for example those that are doing their research in the UK rather than overseas.

Other questions arise around ethics, and the move towards algorithms.

- i. Are you aware of the Government's recent review? Do you think the Government's AI Review will help with these risks and opportunities? Does it go far enough?**

I believe it's a good start, it's extremely light on the commercialisation of AI and the development of the AI market in the UK and many of the peripheral aspects will need to be dealt with such as legal framework, insurance, and dealing with problems such as the misunderstanding of the concept of "explainability", which has now been rather mistakenly enshrined in the GDPR at European-level. This concept completely misunderstands the way decisions are made not just in AI but also by humans and would likely be a major brake on AI development and application in Europe.

Insurance and legal frameworks are just as important catalysts as the more obvious areas, but specific to AI applications.

2. What is the general state of AI start-ups in the UK at the moment, what sorts of challenges do you see them facing at the moment, and how do you think these might change over the coming years?

AI is strong in the UK and there a series of strong start ups. The challenges are those that befall many of our young technology companies – making sure that they have scale up funding and that we have capital markets working so that they can exit so private investors can make a return and they can have access to significant funds. The problems there are generic ones for all technology businesses.

Skills are a key issue. Often in AI companies the employment of many rests on the hyperskills of a few. Those few may not currently be in the UK, so the UK in this field needs to be welcoming and attractive to compete for the global hyperskilled in AI.

For AI, the most specific challenges are access to training data. Frameworks in which further training data becomes available would give them a significant advantage, for example national health information could be made available in the right conditions for these companies, in return for the NHS getting the benefit of the systems created. There should be the concept of strategic data which is data that is of value to the UK and its economy, and this should not be treated as open data. In fact, the academic community in the UK is rather too obsessed with open data, and is in danger of building very large and closed monopolies in other parts of the world which will dominate the market, and this is done through giving our data away without consideration on the basis it is made available.

i. Is the UK an attractive place to invest in AI? What could be done to improve investment in UK AI businesses?

The UK is an attractive place to invest in AI, if problems such as dysfunctional markets for the tech sector can be fixed. There is very good CST report on why the UK stock market failed to grow very large technology business and why they have been acquired by their US competitors

ii. Realistically, can UK companies developing AI systems compete with larger companies in the United States and elsewhere?

There is no reason at all why UK companies cannot compete with AI systems in the US, in fact it has already been the case – Autonomy became a \$11bn business in the UK and it was built on AI. Darktrace a 3year old AI security company is now valued around \$1Bn. The problem is that our bright companies are usually sold to their US counterparts at a value of \$100m. If this was changed and the funding landscape was different in the UK, then these companies could become very large.

Success intorun will create a virtuos circle of skills and investor development.

3. What are the obstacles to AI start-ups scaling up? Is there an investment gap in the UK? If so, how can it be addressed?

The obstacles to AI start-ups scaling are access to data and access to growth capital. The latter is not at the small level of the past of a few million, it's now \$50-\$60m tickets – that really only becomes available when there are such abilities as to list companies and for those to grow in the London markets rather than abroad.

4. Should more be done to prevent the acquisition of UK AI start-ups by larger foreign corporations? If so, what?

It is crucial that young British AI companies are not acquired by larger foreign corporations, as it is our IP that is powering the successes of those organisations. This, however, is best done by making it the more attractive choice for these companies not to sell out and this has to be done by fixing the UK stock market so that these companies have the option to list. At the moment, companies are sold when they get to \$100m because there is no available market in London that functions, so the only alternative is to grow them to \$800-\$900m and take them to Nasdaq. The risk involves in that transition means that many investors choose to sell out, and this needs to be changed.

5. Are there any barriers to collaborating with the higher education sector in order to turn AI research into innovative products? What can be done to foster this collaboration?

i. Should taxpayers expect a return on publicly-funded AI research, and if so, what form should that take?

The biggest in collaboration with the UK in order to create innovative products is the lack of an IP framework, with each university taking a different approach, some of them very unrealistic. The government should put a standard set of IP rules, acknowledging the fact that because the British taxpayer has funded the research, the British taxpayer deserves to see economic impact, benefit and jobs from it. From this, there should be a standard set of terms that produce a frictionless ability to move IP and commercialise it. It should be noted that the universities should not be seeing this as their own private cash cow and in their attempt killing the cow while a calf!

ii. When public datasets are used to develop commercial AI applications, what benefits should the public expect in return?

When public datasets are used to develop commercial AI applications, then any product that results, if bought by a UK government body, should be sold at most favoured nation pricing. For example, if the NHS makes data available that allows someone to create a product that is very useful in the treatment of patients with cancer, then the NHS should get that product at most favoured pricing. This should be explicitly put into the arrangement at the beginning. At the moment, because the UK is dominated by an academic approach, tends to give away its most valuable data and the public services end up paying top dollar when they want to use the service.

6. Do investors have a duty to ensure that AI is developed in an ethical and responsible way? If so, how should they follow through on this?

i. What are your views on the use of regulations or voluntary measures to ensure AI is developed and used in an ethical ways?

In terms of ethics, the issue is to create a legal framework that respects this. There is no ability to create voluntary measures in this area, because there is no agreement and precedent for what is and is not acceptable – there are many open questions and these will be taken in different ways by different people. For example, when we look at analysis of insurance policies, how much taking into account of the data should be allowed when setting the price? AI is going to be very useful but it may also decide things that are unacceptable, such as older people cost more, or people from certain parts of the world are at more risk of certain diseases. So this has to be regulated and it must be wrapped up with not only the constraints, but the opportunities too, so for instance creating legal frameworks and insurance frameworks not unlike those that allowed the airline industry to flourish.

- 7. Should individuals expect to be able to retain ownership of their personal data and still benefit from developments in AI, or are these two aims incompatible?**
 - i. How much control should individuals expect over their personal data? Is enough being done at the moment to facilitate this?**
 - ii. What models of data ownership, management and control would you like to see explored further?**

Obviously the big issue here is that for example if we allowed the treatment profiles – on an anonymized basis – of different treatment options used by the NHS as a data set to an AI, we might well end up saving a lot of lives, money and suffering. If we put in a framework where people own their own data, although this may appear very easy for a lawyer who doesn't understand the field, this becomes a very difficult issue because of derived data – for example, if a machine learning algorithm uses a piece of data in order to learn, then does the owner of the data have an interest in the algorithm? In reality, what has to happen here is a sensible set of rules need to be put together to allow this area to develop the benefits. This is one where probably it is better for people to be able to actively opt-out rather than opt-in, that way sufficient data will be generated to get the benefits, for example in healthcare, but people still have the option of their data not being used.

- 8. To what extent do you expect established techniques, like deep learning, to continue delivering meaningful progress in artificial intelligence?**
 - i. Could there be limitations to these approaches, and if so, what might be done to overcome them?**

One could certainly have a debate as to whether deep learning is an established technique! The question is not about a particular algorithm; this is an area where the methodologies in the most basic sense of data-driven probabilistic self-learning systems will unfold in many different ways as research continues. There are limitations but often these are more fundamental to the problem than the solution, meaning that whilst we like to think that everything has a correct answer, most of the real world problems don't have one correct answer, they have probabilistic answers and so the first thing we need to do is understand that – it is this kind of misunderstanding about how the world works before we really even get to AI that has led to the major mistakes made in the GDPR explainability rules, which will be a great hindrance to the sector.

- ii. Do more experimental, unusual or radical ideas in AI deserve more investment? Are there opportunities for private investment, or**

should the Government take a more active role in supporting these areas?

Probably not – one of the big dangers of AI is that one can produce a demonstration without understanding why it works and when it works and so actually it very important that the field is encouraged to work on sound basis – we have seen this in the 80s with neural network development so I think it very important that excellence and rigour are maintained as criteria for funding research in the area, rather than looking for things that are radically different. Obviously, though, there can be things that are different that are based on a rigorous approach. The government already supports AI research well in the UK, what it needs to do now is support things that more actively bring it together especially application areas such as the Turing Institute.

10 November 2017

**Ms Nika Mahnič and Professor Kathleen Richardson –
Written evidence (AIC0200)**

[Submission to be found under Professor Kathleen Richardson](#)

The Market Research Society – Written evidence (AIC0130)

Background: About the Market Research Society (MRS) and the research market

1. The Market Research Society (MRS) is the world's largest research association. It's for everyone with professional equity in market, social and opinion research and in business intelligence, market analysis, customer insight and consultancy. MRS has 5,000 members in over 50 countries and has a diverse membership of individual researchers within agencies, independent consultancies, client-side organisations, the public sector and the academic community.
2. MRS also represents over 500 research service suppliers including large businesses and SMEs plus a range of research teams within large brands such as Tesco, BT, ITV, Telefonica and Unilever which are accredited as MRS Company Partners.
3. MRS promotes, develops, supports and regulates standards and innovation across market, opinion and social research and data analytics. MRS regulates research ethics and standards via its Code of Conduct. All individual members and Company Partners agree to regulatory compliance via the MRS Code of Conduct and its associated disciplinary and complaint mechanisms.
4. The UK is the second largest research market in the world, second to the US, and in terms of research spend per head of population is the largest sector with £61 per capita in 2015 (with the US at £39, Germany £24 and France £23)⁸⁶⁶. The UK research supply industry is a £4bn market and has grown steadily over the previous five years by an average of 6% per year⁸⁶⁷. In 2016, MRS with PWC undertook an updated assessment of the size and impact of the UK research and evidence market, *The Business of Evidence 2016*⁸⁶⁸. One of the main findings from this report is the size of the UK 'business of evidence' market, which employs up to 73,000 people and generates £4.8 billion in annual gross value added (GVA). Data analytics exhibits the highest growth rate at over 350% growth since 2012.

⁸⁶⁶ See the Research-Live Industry Report 2017:

http://www.mrs.org.uk/pdf/MRS_RESEARCH%20LIVE%20REPORT%202017%20.pdf

⁸⁶⁷ See ONS Annual Business Survey:

<https://www.ons.gov.uk/businessindustryandtrade/business/businessservices/bulletins/uknonfinancialbusinesseconomy/2015revisedresults>

⁸⁶⁸ See Summary of Business of Evidence report 2016 at https://www.mrs.org.uk/pdf/boe_info.pdf.

5. The UK research sector is recognised as leading the way in the development of creative and innovative research approaches including maximising the opportunities afforded by the development of new digital technologies. The methodological issues are explored and debated in the academic journal, the International Journal of Market Research.

Submission: Ethics and Role of Government

6. MRS welcomes the opportunity to respond to the Call for Evidence by the Select Committee considering the economic, ethical and social implications of advances in artificial intelligence. Our response focuses on the ethical implications of artificial intelligence and suggests that the regulatory framework needs to encompass both legal and self-regulatory initiatives that build consumer trust.
7. Market research, which includes social and opinion research, is the systematic gathering and interpretation of information about individuals or organisations using the statistical and analytical methods and techniques of the applied social sciences to gain insight or support decision making. Research itself does not seek to change or influence opinions or behaviour. Artificial intelligence (AI) with its complex software systems is used in the research sector especially in data mining and analysis. In big data analytics developments in natural language processing (NLP) and machine learning models are aiding the automation of data analysis, data collection and report publication.
8. The General Data Protection Regulation, which will come into force in the UK on 25th May 201, provides a robust legal background for control of artificial intelligence with strengthened individual rights, a focus on transparency and accountability and provisions that address automated decision making. However this framework needs to be underpinned by continued regulatory guidance from the Information Commissioner's Office (ICO) such as their recent carefully balanced informative paper "Big data, artificial intelligence, machine learning and data protection" which carefully considers key issues such as privacy and consent.⁸⁶⁹
9. Critically, in order to control AI, the regulatory framework will need to evolve nimbly and flexibly with respect for protection of personal data of individuals at the core. In light of this the legal framework will need to be supplemented by an ethical framework – based on by self-regulatory and trust frameworks such as such as the MRS Code of Conduct and the MRS Fair Data Scheme -

⁸⁶⁹ICO Big Data Paper:430

<https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

which will ensure that organisations take their obligations seriously and implement AI and use data in a fair and transparent manner.

10. Consumer facing marks with consumer recognition are a useful tool in building consumer trust across markets and can form a vital part of the framework for regulating the use of AI. Organisations, public, private and not-for-profit, using AI need to understand the nuances of consumer privacy preferences as it applies to their particular market and organisation in order for them to properly determine the right balance in the use of AI. Use of ethics boards and ethics reviews committees and processes within a self-regulatory framework will be important tools.
11. MRS adopted its first self-regulatory Code in 1954 and the latest fully revised version of the MRS Code of Conduct came into effect on 1 September 2014.⁸⁷⁰ It is currently being revised to reflect legal changes and ensure it reflects emerging technological developments. Supported by a suite of guidance documents the Code supports those engaged in market research in maintaining professional standards and reassures the general public that research is carried out in a professional and ethical manner. The MRS Code is technology and methodology neutral. It sets out overarching ethical principles supported by rules of conduct.
12. Additionally, a broader range of firms have signed up for the MRS Fair Data mark, which was established in 2012 to complement the self-regulatory arrangements under the Code. This trust mark, is designed for use by consumer-facing firms, suppliers of research and data services, and public bodies. It enables consumers and citizens to make educated choices about their data and to identify organisations which they can safely interact with, knowing that their personal information is safe. For organisations that are accredited, it demonstrates a commitment to be ethical, transparent and responsible with data. Organisations sign up to ten clear principles that are consumer focused, enabling ease of understanding. These ten core principles of Fair Data work in tandem with the MRS Code of Conduct. The scheme is supplemented by MRS's Fair Data advisory service which includes face-to-face, telephone and e-mail support plus events on best practice, best practice guidance and a bespoke audit accreditation process which is mandatory for all organisations that are not MRS Company Partner accredited.
13. We believe data privacy, compliance and ultimately building consumer trust, are critical. If the public become more afraid of sharing their personal data, the effect could have long term implications for research participation, development of innovative commercial solutions and society. This is even more critical in light of the increased use of AI where opacity of processing

⁸⁷⁰ MRS Code of Conduct <https://www.mrs.org.uk/pdf/mrs%20code%20of%20conduct%202014.pdf>

and varied levels of human intervention raise specific concerns for individuals. Credible, robust self-regulation and trust marks will be key tools in raising awareness about the collection and use of data and assist both firms and consumers in benefitting from the use of AI. Codes of conduct can adequately address these issues, using consultative processes, to ensure codes enshrine privacy and transparency and reflect societal expectations from use of this new constantly evolving technology.

6 September 2017

Professor James Marshall, Professor Thomas Nowotny, Dr Andrew Philippides and Dr Paul Graham – Written evidence (AIC0088)

Submission to House of Lords Select Committee on Artificial Intelligence

Professor James A. R. Marshall, Department of Computer Science, University of Sheffield

Professor Thomas Nowotny, Department of Informatics, University of Sussex

Dr Andrew Philippides, Department of Informatics, University of Sussex

Dr Paul Graham, School of Life Sciences, University of Sussex

September 5th 2017

Introduction

1. We write as investigators on a substantial 5-year EPSRC Programme Grant entitled 'Brains on Board'^{871,872}, which seeks to reverse engineer the honeybee brain in order to develop autonomous adaptive controllers for unmanned flying vehicles, among other applications.

2. The definition of 'artificial intelligence' we work under is that of reproducing animal-level autonomy and learning abilities in computational form. Our usage is close to the original definition of artificial intelligence during its emergence in the 1950s as the computational description and simulation of biological learning and intelligence, and our aims are also close to the original in that our simulations should both replicate animal-like capabilities in a robot and shed light on the biological basis of intelligence.

3. This definition, our aims and thus our research, is distinct from 'machine learning' approaches to intelligent behaviour, such as 'deep learning' applied to image recognition or the playing of video games, where the underlying algorithm is not a simulation of the diverse processes generating those behaviours in real brains. However, a common thread across much AI and robotics research is that developed solutions, algorithms and robots are designed for a specific purpose, task or 'ecological' niche. This distinction between AI as originally imagined and the narrower use of AI as machine learning or data science runs through our responses.

4. Our responses are also informed by the fact that as Investigators on this large EPSRC investment, we also represent over 100 years of collective

⁸⁷¹ <http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/P006094/1>

⁸⁷² <http://www.brainsonboard.co.uk/>

experience of higher education and the UK research landscape, and have witnessed first-hand the transformation in AI usage and teaching over the past two decades. The applications of modern AI are likely to be widespread across academic, research and technology sectors; this has implications for the training and skills required by STEM students, especially those outside of Computer Science.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

5. The current state of AI reflects recent developments in hardware, computational power and 'big data', that have caught up with the potential of techniques developed over the previous decades. Machine learning techniques can now to be applied to hard problems (Go⁸⁷³, Atari games⁸⁷⁴, Jeopardy!,⁸⁷⁵ driving a car) with impressive success. These successes are based on algorithms utilising large amounts of specific data and computational power, whereas humans, for instance, do not require millions of games of Atari to be able to play. Therefore, we are not dealing with examples of general purpose human-like, or even animal-like, intelligence. Furthermore, algorithms are restricted to narrow niches, show minimal generalisation, and are fragile; for example, Tesla's self-driving mode could not deal with traffic crossing in front of the car well, whereas humans can generalise driving skill in terms of obstacle avoidance.⁸⁷⁶

6. Even without further technological advances, we expect development in AI to be significant in the next 10+ years, principally through the creative application of current AI technology to new products and in new research situations. This may be hindered by a lack of skilled individuals, through over-zealous regulation, or through public mis-trust.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

⁸⁷³ Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.

⁸⁷⁴ Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.

⁸⁷⁵ <https://www.scientificamerican.com/article/deadly-tesla-crash-exposes-confusion-over-automated-driving/>

⁸⁷⁶ <http://www.bbc.co.uk/news/technology-30290540>

7. Excitement has two connotations, both positive and negative. Negative views have been advanced by prominent public scientific figures, such as Stephen Hawking⁶ and Elon Musk⁸⁷⁷ claiming that AI poses an existential threat to humanity in the long term. In the shorter term, reports have claimed that 10m jobs in the UK are under threat from AI within the next 15 years.⁸⁷⁸ At the same time, anticipation of the economic and societal benefits of AI in the short to medium term has been mounting over recent years; analysts suggest AI may enable growth in economic productivity of between 10% to 100% in the UK, as in other developed economies, over the next 15 to 20 years.^{879,880}

8. There is precedent in the current excitement generated by AI advances. As mentioned in our introduction (paragraph 2), AI as a recognisable field originated in the 1950s. Following this, and the public interest generated, a great deal of anticipation developed. The AI methods of the day differed substantially from the successful technologies in use today, being based largely on the idea that brains could be treated as symbol-manipulating computers, but the current promises of an AI revolution are very reminiscent of expectations decades ago. Ultimately these expectations proved to be grossly inflated, resulting in the famous 1973 Lighthill report⁸⁸¹ and the subsequent restriction of AI funding within the UK referred to as the 'AI Winter'.

9. Given the limitations of current AI technology discussed in answer to q1 (paragraph 5), there is a risk of over-promising and under-delivering. However, even given this, it seems unlikely, given substantial government and commercial funding, and the commercial and potentially societal advantage that comes from application of machine learning to real-world datasets, that there will be a second AI winter.

10. Following our response to q1 (paragraph 5), we conclude that certainly fears of an existential threat to humans from AI are not warranted in the foreseeable future, based on current technologies. However, excitement around the economic opportunities presented by AI seems likely to be justified and, in fact, self-fulfilling, due to the level of investment by the major technology companies, currently and for the foreseeable future. With a well-established research base in AI, and related fields such as robotics and data science, the UK

⁸⁷⁷ <https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat>

⁸⁷⁸ <https://www.pwc.co.uk/services/economics-policy/insights/uk-economic-outlook/march-2017.html>

⁸⁷⁹ <http://www.pwc.co.uk/services/economics-policy/insights/the-impact-of-artificial-intelligence-on-the-uk-economy.html>

⁸⁸⁰ <https://www.accenture.com/gb-en/insight-artificial-intelligence-future-growth>

⁸⁸¹ Lighthill, J. (1973): "Artificial Intelligence: A General Survey" in Artificial Intelligence: a paper symposium, Science Research Council

is well-placed to see a direct economic benefit from this investment; for example UK start-up Deep Mind attracted a reported \$400m of investment from Google in 2014.⁸⁸² It should be noted, however, that while the UK's expertise in artificial intelligence is recognised, the size of government and private investment in development of AI technologies within the UK is dwarfed by that in the US.⁸⁸³

11. We do not explicitly comment on the potential employment or economic impact of machine learning here as, economically, the two are likely to be hard to distinguish.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

12. Given our response to q2, we feel that AI practitioners are morally obliged to communicate effectively with the public on the true nature of current AI technology and its likely development within the next 5 to 50 years. For instance, we feel it is unlikely that the general public are aware of the extent of AI in the guise of machine learning and data science in existing day-to-day technologies (e.g. face recognition in Facebook, voice recognition in SIRI, Google Translate, pattern recognition in spending habits), some of which utilise large personal datasets. Public engagement can help address this knowledge gap. We note that only 1 exhibit in the last 4 years of Royal Society Summer Science Festivals (~100 exhibits in that time) concerned AI or machine learning.

13. Educationally, computer science and the applications thereof need to be integrated more fundamentally into school and university curricula. As a skillset, coding and algorithmic thinking are lacking, even in academia, and should not be seen only as a necessary skill for computer science. Academia, engineering, biological and medical sciences, broad research, technology and finance sectors will all depend on skilled people comfortable with the implementation, or understanding of AI and machine learning algorithms.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

⁸⁸² <https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind>

⁸⁸³ <https://techcrunch.com/2015/12/25/investing-in-artificial-intelligence/>

14. As noted in our response to q2, the development of artificial intelligence promises significant new economic opportunities, but these will be disruptive, or even destructive. Fears of job losses are warranted, and will differ by sector and by sector uptake. Driverless car technology, for example, poses a very real threat to professional drivers in the taxi and haulage industries; likewise, much of the development in robotic automation of engineering processes and farming are aimed at reducing the amount of human labour required; even comparably low uptake in such sectors would lead to substantial redundancies. Excitement around the economic activity arising from development of AI technologies is likely to be justified, given the history of the most recent disruptive technology, the internet; quantitatively predicting the economic value of such activity seems susceptible to serious inaccuracy, but observing the current market capitalisation of the major internet technology companies is likely to be indicative.

15. The challenge for government will be that while economic activity in AI will generate employment and revenue, this will most likely be in small numbers of highly skilled jobs; however jobs lost will be in larger numbers, and will be comparatively unskilled. Increasing the UK skills base in technology will thus only be a partial solution. This is an obvious economic redistribution that must be anticipated if it is not to lead to societal imbalance and its attendant problems, as has been witnessed in the UK before, most recently in its industrial towns and cities.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

16. See our response to q3

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

17. There are clearly important ethical issues around the development and use of AI. We believe that one ethical principle for AI has to be that the AI itself is not an ethical agent, similar to the principles formulated for robotics.^{884,885} The

⁸⁸⁴ BS8611 Robots and Robotic Devices. Guide to the ethical design and application of robots and robotic systems. British Standards Institute

⁸⁸⁵

<https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>

developers and operators of AI systems are the true ethical agents. In terms of ensuring ethical, legal, safe and unbiased behaviour, the most immediate measure is to require the same standards from an AI system that would be required for systems operated by humans. For example, a self-driving vehicle would need to pass the same safety tests as a normal vehicle and demonstrate the same road safety competency as a human driver as judged in driving tests. For AI systems the test may have to be more extensive because the common assumption that, if a human can operate a car safely in a given half hour period, she or he will be able to do so generally, may not hold for an artificial system. A self-driving car should therefore be tested under a multitude of situations and conditions. But the principle of judging the system based on its performance in the task it is designed for seems sound. This indicates that certification of AI systems is likely to be behaviour-based and statistical, for example in terms of numbers of failures per million miles driven as in the case of driverless vehicles, rather than based on logic-based systems verification (see also our response to q9, paragraph 18).

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

18. AI in the sense we defined above will necessarily have a black box nature as complex algorithms exhibit 'explanatory opacity', not least because one may not be able to accurately capture the inputs (ground/weather conditions) that may have caused any particular behaviour. This is similar to humans where the true underlying mechanisms that lead to any given action or decision are not known. The same is also true for many systems in machine learning, e.g. for all deep learning networks, and even for complex software systems involving no learning at all. One can nevertheless require some transparency on what input data was used to arrive at an output/decision/action and research into 'transparent machine learning' of this kind is under way.⁸⁸⁶

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

19. From our position as part of the University system and the UK research community, we would like to see policy and investment aimed at equipping future generations with a skill-set fit for the future. School curricula, UK degree portfolios and interdisciplinary research funding can all be considered,

⁸⁸⁶ <http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/P03442X/1>

representing the reality that all sectors will be influenced by AI, machine learning, and related disciplines such as robotics. (see response to q3, paragraph 13).

20. Such investment would facilitate access to the new economic sectors regardless of socio-economic background, as well as strengthening the UK's capacity for further growth in artificial intelligence and related fields. However, as noted in our response to q4 (paragraph 15), a systemic consideration of how to compensate for the potential loss of large numbers of unskilled jobs within the economy is clearly the sole responsibility of government. We anticipate that other evidence submitted to this select committee will also identify this risk, and propose concrete policy measures that might be taken.

21. In terms of regulation, the UK has an opportunity to lead in translating the emerging principles of responsible innovation in AI and related disciplines, into a regulatory framework; as discussed in our response to q8 (paragraph 17) these include assigning agency and responsibility to humans and organisations, rather than to AI systems, and establishing methodologies for certifying that systems we do not fully understand, with complex emergent behaviour, are safe for their intended purpose.

22. We assume that other experts will have submitted evidence on other regulatory issues such as establishing principles for insurance of losses resulting from the actions of AI-based systems.

5 September 2017

Dr Neil McBride – Written evidence (AIC0047)

The ACTIVE Ethics Implications of Artificial Intelligence.

Neil McBride PhD

Centre for Computing and Social Responsibility
De Montfort University

- 1.1 This submission develops the ethical implications of artificial intelligence using the ACTIVE ethics framework which addresses the issues under the headings of Autonomy, Community, Transparency, Identity, Value and Empathy. Artificial intelligence is the use of human artefacts to collect, collate, analyses, draw conclusions and make decisions based on data gathered from the informational and physical environment.
 - 1.2 It is critically important to consider AI from a human-centred point of view. It is the interaction between the machine and the human, the machine and community, the machine and society which most affects the ethical outcome of AI use.
 - 1.3 The following considers each of the six facets of ACTIVE ethics and addresses some of the questions raised by the Select Committee on Artificial Intelligence.
- 2.0 Autonomy: To what extent is the user master of his own information and in control of his interactions with the AI?
 - 2.1 The totally autonomous car, for example, will take all its information from an analysis of its environment without reference to humans or an external information input. However, total autonomy is not sustainable since all AI requires human input and connection. Humans should not be eliminated from systems and roles by their replacement with AI. Rather AI needs to enhance and complement human skills and abilities.
 - 2.2 Increasing the autonomy of the machine may reduce the autonomy of the human who is disempowered and rendered a passive recipient of decisions made by the machine. It is critical that the autonomy is shared between human and machine and that the user retains some responsibility.
 - 2.3 The removal of risk through the use of autonomous decision making machines shifts responsibility away from the user and may result in rash user actions which actually decrease safety and increase risk.

2.4 There is an urgent need for the education of the public in the nature of autonomy, what is meant by, for example, the term autonomous car, and how power is redistributed by the use of AI.

2.5 The widespread use of AI will rearrange boundaries of responsibility. For any implementation of AI there must be clear statements of the boundaries of responsibility and the allocation of both ethical and legal responsibility between stakeholders.

3.0 Community: How does the AI support and develop the community within which it resides?

3.1 AI supports communities. Its job is to enhance and support individual and community capabilities. It does not operate on its own, isolated from human and society, but is part of a team, connecting with humans and supporting the free will decisions of humans.

3.2 AI in communities should support the humans and communities in enabling individuals, groups and societies to learn from mistakes. It should support the distribution of resources to fulfil the common good, and should be judged on the importance the AI attaches to children, the disabled and disempowered groups.

3.3 Decisions about the deployment of AI should be located in the communities, subjected to discussion, and subject to democratic processes.

3.4 AI has the potential to centralise power and resources. Power may be centralised in the hand of experts who can understand and manipulate the algorithms and corporations who have the resources to invest in development and manufacture. In this scenario the community may become disenfranchised.

3.5 The response of the disenfranchised community may be one of passive acceptance or resistance catalysed by fear and lack of understanding. Locally deployed implementations of centralised AI systems such as smart traffic systems and smart meters should be supervised and regulated by local democratic bodies.

3.6 An AI equivalent of the Food Standards Agency, the Autonomous Systems Standards Agency, monitored locally and distributed by the equivalent of an environmental health officer.

4.0 Transparency: Is the derivation and use of the AI clear to the users?

- 4.1 Transparency is essential for ethical deployment of AI. But this has to be tempered by the need for knowledge to understand how, for example, particular algorithms or robotic deployments work.
 - 4.2 There should be no attempt to create illusions in which the AI appears to imitate human characteristics without the human user being aware that this is happening.
 - 4.3 While technical understanding of a black box may not be completely necessary, information concerning the limits and boundaries of the AI capabilities should be explicitly stated. For example, in the use of Personal Health Monitoring systems the assumptions involved in the algorithms, the limits of accuracy and possible error conditions should be stated.
 - 4.4 In any public sector deployment of AI, contracts, project plans, records of deployment must be available in the public domain. Legal mechanisms must be in place to challenge any 'commercial in confidence' arguments for obscuring or hiding information on the capabilities, algorithms and limits of any AI deployed within the public sector.
 - 4.5 Provenance, origin, and basic algorithmic strategy must be available and freely accessible.
- 5.0 Identity: How does the information system affect the user's identity and purpose?
- 5.1 Identity concerns a person's concept of who they are, the moral and social beliefs they embrace and how they relate to others. The deployment of AI, for example to analyse genomes for the possibility of disease may change a person's identity.
 - 5.2 Big data, which include information the individual sheds in her daily life, health records, government records, loyalty cards and so on render information privacy impossible. The individual cannot possibly know and control where all the information about herself is, as this is distributed across many systems including social computing systems and is in many forms.
 - 5.3 AI enables automated decisions which affect the individual and her identity to be made quasi-autonomously. Those decisions may disrupt or change the individual's perceived identity.
 - 5.4 Although it is not possible for the individual to control the flow of information about herself, nor AI decisions resulting from the activity of algorithms, the individual can maintain control of the implementation of

those decisions. The individual must be able to refuse action based on the AI-generated decision which may affect her perceived identity.

5.5 Not only can the outcome of AI decision making disrupt a person's identity, but it can also create a person's identity.

6.0 Value: How can the value of the human as an individual and a member of society to be respected when interacting with AI?

6.1 AI risks devaluing the human and reducing him to a package of information which defines him as an individual and which can be manipulated as big data, or as part of a smart system. The information is not the person, but just an interpretation of the person, Therefore, any attempt to exclusively control or guide the person using information may result in a drift from reality and the creation of an impression that the individual is the information.

6.2 The value placed on information, and hence the protection given to the information must be interpreted in terms of the value of the subject of the information. Information bought and sold without taking into account the subject might be viewed as a kind of slavery.

6.3 But information is predominantly not about the individual, but about relationships whether with other individuals or organisations. Information tracks relationships. Hence the value of AI lies in the human relationships it supports. Therefore it is important to understand the relationship supported by AI and legislate and regulate to set boundaries for or to manage the relationships. We will need to understand how AI mediates and changes relationships.

6.4 The value of AI should not solely be judged in terms of external measures of efficiency, cost saving and profit. Rather we need measures of internal good and human flourishing. These will require a clear understanding of the human purpose of the AI. Measures addressing intensity, adventure, extension of capabilities, development of courage and human engagement will need to be developed. This will require new research into how to measure the human effect of AI.

7.0 Empathy: Does the information systems professional understand the effect of the AI on the user and their tasks?

7.1 AI developers, manufacturers and legislators need to put themselves in the user's and public's shoes. How do they feel about the presence of a robot on the street? What does it feel like to be moved round in a driverless car? Personal fears and apprehension of AI should not be dismissed but explored and taken into account. Empathy is at the heart

of an Ethics of Care which involves the learning of compassion and benevolence.

7.2 The development of empathy will enable the effective deployment of AI in ways which are appropriate and beneficial and will mitigate the possibility of the public rejection of AI.

7.3 Professional training standards will need to be developed for AI engineers and practitioners. A service-based concept of AI engineering will be required which views the manufactured AI software or artefact in the context of a service delivered and the interaction with the human customers in the achieving of a goal.

7.4 A training framework will be required which turns AI practitioners into reflective practitioners who are aware of the human environment within which the AI acts. For example, a 4Rs framework might be applied in which an outward Reconnoitring of the environment is undertaken; the practitioner Realises the social impact and effects of the AI, Reflects on the changes that might be needed and Revises the design of algorithms and the processes and plans for implementation.

8.0 Conclusion

8.1 The rise of recent rise in popular interest in AI is due to the increase in computing power and social changes rather than any new technical discoveries.

8.2 Education and transparency are essential for enabling the widespread use of AI.

8.3 An understanding of how autonomy and how AI alters power balances in society should be developed.

9.0 Bibliography

Neil Kenneth McBride , (2014), "ACTIVE ethics: an information systems ethics for the internet age", Journal of Information, Communication and Ethics in Society, Vol. 12 Iss 1 pp. 21 – 44

Neil McBride and Robert Hoffman (2016) Bridging the Ethical Gap: From Human Principles to Robot Instructions. IEEE Intelligent Systems. Sept / Oct 2016 76-82

Neil McBride and Bernd Stahl (2014) Developing Responsible Research and Innovation for Robotics. Proceedings of the IEEE 2014 International Symposium on Ethics in Engineering, Science, and Technology

1 September 2017

Mr John McNamara – Written evidence (AIC0081)

SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

Call for evidence

John McNamara

Senior Inventor

Please note this is a personal submission

Definition of A.I

There are a few definitions to choose from that hit the mark fairly well:

“...hardware/software that mimics the function of the human brain and helps improve human decision making” – Wikipedia

“...systems that learn at scale, reason with purpose, and interact with humans naturally” – IBM

And my own definition:

“A technology which successfully executes tasks, that generally one would consider only a biological intelligence would be capable of performing.” – John McNamara

Pace of Change

What is the current state of A.I and what factors have contributed to this?

To answer this question we need a bit of context.

Rather than a brand new phenomena A.I is really the extension of a technological change that started over two centuries ago with the industrial revolution. This industrial revolution is considered by many to be the “1st Machine Age” – which had the effect of scaling human physical capacity massively. So, for example rather than 20 men being required to move a log into a factory, by hand, you could have 1 man move the log into the factory, using a steam powered pulley/crane.

With the popular advent of A.I, many are considering today to be the “2nd Machine Age”. This Machine Learning Age, will scale human mental capacity, to the same degree (if not more) than the industrial revolution scaled human physical capacity.

The current state of A.I today, is that we are growing and perhaps transitioning (not always successfully) from a top-down rule based intelligence, (with human written rules that are easy to understand, investigate, fix if necessary) to a bottom-up Deep Learning intelligence which utilises several layers of artificial neural networks, that to a degree, mimic the way the human brain works. These neural network are 'trained' on data, and are then able to identify patterns in the information provided (identify a car, or a person in the road). The problem with this adopting a purely a Deep Learning approach is that (much like the human brain it is modelled on) once trained you cannot take it apart to understand the 'rules' it uses to identify these patterns. These problems manifest in A.I often mistaking correlation for causation – effectively introducing bias into systems (there have been some very disturbing results when neural networks have been applied to profiling potential criminal suspects, in terms of race/ethnicity)

This means that although we are seeing these Deep Learning neural networks being applied to critical systems (autonomous cars for example) we don't really understand fully why they act the way they act – we very often see results that appear to be correct and assume that the network is making the right decisions. This means that as of today (in my opinion), we cannot depend upon purely neural network intelligent systems, if we are required to analyse how, and why a conclusion has been reached. This is the current quandary, which is leading to A.I systems which are a melding of top-down and bottom-up A.I technologies

How will A.I develop over the next 5, 10, 20 years

Five years

Much like the availability of power during the industrial revolution enabled machines that were once powered by manual labour, to be powered by steam or electricity, we (I believe) will see a huge growth in start-ups/businesses who's aim is to transform everyday household, industrial and business technologies that are 'dumb' and apply A.I to them. This will transform the way 'dumb' technologies interact with humans, so for example providing your car with visual recognition so it recognise you and can adjust your seating, climate control automatically. Or perhaps coffee machines, recognising you and creating your coffee just the way you like it. We will see technologies react to the way you look 'happy/sad/pleased/displeased' and act accordingly. We will certainly see a trend to create more Assisted Living technologies that interact with the elderly through voice and visual recognition.

As computing power increases, and so the number crunching ability of A.I increases we will see advances in scientific/medical research. For example let's take health research, as patterns are found in the myriad of research papers that the medical profession creates (and are impossible for a single researcher to keep abreast of). The most relevant patterns and insights found in these papers will be automatically provided to the researchers of new treatments, providing

them with timely insight and enabling them to execute and progress more rapidly.

Ten years

Huge disruption in service industry & retail industry will occur. At the moment, rudimentary A.I is being used to service customers who know exactly what they need. In the next ten years, A.I systems will be able to recognise customers, & will have access to their habits, styles, personality traits (via social media – or customer database) and be able to make ultra-tailored proactive recommendations/suggestions to their users (up sell/cross sell). This will cause societal disruption, as unless managed this trend will cause widespread unemployment and a social group who will be socially disenfranchised. We will also see disruption in the way that our services are consumed. Rather than personal interaction we will start to see the beginnings of our own A.I Digital Avatars. These avatars will be A.I's which will be governed by our own tastes, likes, dislikes, political views, and modelled on our own personal interaction styles. These will be used to conduct day to day management of more menial aspects of our lives (from an avatar skilled in utility vendors, which will find the right electricity supplier and switch on your behalf, or an avatar skilled in auto-mechanics, dealing with your next car service based on your driving style, to a nutritional avatar who deciding what your weekly meals should be and orders/creates them). We will also see the application of these avatars to areas such as politics – where we will instil in our own personal Political Avatar, our political leanings and have it scour all available data (from Hansard to the Daily Mail) to provide you with a recommendation on who to vote for and why, based on your world view.

20 years

We will see the miniaturisation of A.I systems so that intelligence is ubiquitous in our environment. Also we may see A.I nano-machines being injected into our bodies. These will provide huge medical benefits, such as being able to repair damage to cells, muscles and bones – perhaps even augment them. Beyond this, utilising technology which is already being explored today – we see the creation of technology that can meld the biological with the technological, and so be able to enhance human cognitive capability directly, potentially offering greatly improved mental, as well as being able to utilise vast quantities of computing power to augment our own thought processes. Using this technology, embedded in ourselves and in our surroundings, we will begin to be able to control our environment with thought and gestures alone. Unfortunately, the risk is that this capability will come at a (monetary) price. If our technological reach exceeds our more materialistic grasp, whereas today, being poor means being unable to afford the latest smart phone, tomorrow this could mean the difference between one group of people potentially having an extraordinary uplift in

physical ability, cognitive ability, health, life span and another much wider group that do not.

Is the current level of excitement which surrounds A.I warranted?

I think it depends on who you ask. From a business, research and technology perspective, the excitement is definitely warranted. It offers the opportunity for unparalleled disruption in current services and the creation of wholly new industries. The most enlightened business leaders will use this technology for growth, new services and facilities and those that can afford to avail themselves of these services should be very excited. Unfortunately less imaginative business leaders will only see adoption of A.I as a route to rapid cost reduction in staff.

Much like splitting the atom, there is unparalleled opportunity for the technology to better society, but there is also the opportunity to damage our society irreparably. I recommend that the effects of widespread application of A.I be considered and potentially managed, before widespread A.I application to industry is implemented.

Ethics

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

I recently spoke about this at the NeSy17 conference in an Industry Panel debate with Google and Accenture - <http://bit.ly/nesy17JM>

For me this is quite a simple question.

Any business system where there is a requirement or need to be subject to an audit, must not implement an A.I system that has a 'black box' element. This would render a large portion of the justification/reasoning behind business actions that are subject to such an audit, opaque.

Any A.I system who's findings could impact to alteration of legal status for a citizen must not include a 'black box' A.I algorithm. All reasoning and decisions that affect the legal status of a citizen must be completely transparent and open to question.

Any A.I system that could potentially be the cause or participate in the harm of a citizen must be transparent, auditable and not subject to 'black box' decision making. This is currently topical in the implementation of A.I in autonomous cars.

For more less critical uses of A.I, for example in assisting a user in coming to a decision, the use of black box A.I should be permissible, but only so long as the

actual decision is made by a human being, and that they are comfortable that they understand the rationale behind the decision. “Computer says ‘no’” cannot be a justification for a decision. An A.I system must be an augment to decision making, not the decision maker.

The Role of Government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

The only real way that Gov’t can have a meaningful role in the development and use of A.I, is to initiate projects that use A.I (perhaps to solve societal problems). By being in the forefront of the research, the Gov’t will be in a position to understand the benefits and the dangers of A.I.

If the Gov’t were to take a more re-active posture to A.I development and implementation, I fear that it will always be in a position of closing the barn door once the horse has bolted. It will be focused on treating the societal and economic symptoms of a problem (which will be painful and expensive) of unwise, short term benefit applications of A.I, rather than being in a position to identify and steer away from any such pitfalls.

My suggestion here would be to initiate a series of A.I ‘Apollo’ like projects, where we utilise technologists to research and implement solutions to a particular societal problem (let’s say improving education) with A.I. Not only would there be benefit in solving the problem it was designed to solve, but the tangential technologies and industries that would be created would be a huge financial boon to the Gov’t and the UK could be seen as a world leader in a technology that would undoubtedly become increasingly more crucial over the coming years.

In terms of more general regulation, there should be a generally accepted ‘ethical standard’ which is embedded into A.I systems. Any application of A.I to a system over and above those systems that simply augment human decision making, should be subject to a regular audit, to ensure that the rules that govern the ethics of that system are comprehensible, correct, coherent and are aligned with the ethical design of the system. This could range from a McAfee style software that interrogates A.I systems in smaller applications, to manual audits led by technologists in more business/human safety critical applications.

5 September 2017

Sabine McNeill – Written evidence (AIC0009)

- 1) Whilst I respect the necessity for anonymization, I am writing from the unusual experience of one of the first three women in mathematics and computing at CERN, where I worked as a systems analyst and software diagnostician. I bought my first Apple in 1979 and later re-visited maths through the eyes of a mature programmer, practising 'software-aided thinking' in 'splendid intellectual isolation'. My insights were astounding regarding the history of science and the confluence of coding and maths as an expression of our thinking, with CODING as a new technical skill and DATA a new digital asset.
- 2) However, having left CERN to become self-employed, I did not publish my understanding but wrote and eventually designed software that encapsulates my thinking. An American IP lawyer advised me that 'blackboxing' is my best protection as a 'trade secret', in a world where patents are the game of the 'big boys' and only buy the right to defend oneself after violations.
- 3) The outcome as an SME struggling for funding is Smart Knowledge – work in progress that combines the best of humans, pattern recognition, with the best of machines: number crunching. The generic nature of this system visualises IMAGES, MULTI-DIMENSIONAL DATA and TIME SERIES in new ways. Independent of scale and of application, it effectively unites time and space on screen. It also accepts all imaging technologies, such as microscopes, x-rays, cameras and telescopes, besides any data from financial markets and climate change to medical and environmental applications.
- 4) Hence I am advocating its use as a way to address not only the gap of data skills in the UK, but to develop scientific world views based on seeing data through software lenses. Its use will also develop new industry standards and scientific references by seeing more depth, detail and structure in images, and dependencies, relationships and priorities in data, besides automating quality and process control.
- 5) This unique solution is a very strong antidote to all negative implications of AI. Hence I hope it will be welcomed by this timely consultation. But maybe 'digital' and the pervasive power of the internet have already outpowered all noble attempts to control the development of AI?

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- 1.1. Did you Know? Shift Happens! was created in 2006 by US educators to illustrate the pace of technological change without being able to keep up-to-date – but making the point that technology itself determines the pace, e.g. the world wide web and mobile phones.
 - 1.2. AI has to be seen as an extension of institutionalised thinking that is beyond anybody's control – but a consequence of the economics of 'winner takes all' and the corporations that are 'too big to fail' – beyond anybody's accountability or governance.
 - 1.3. Calls to add A for Arts into STEM [Science, Technology, Engineering and Maths] are a sign of the necessity to change direction.
2. Is the current level of excitement which surrounds artificial intelligence warranted?
- 2.1. Just as beauty lies in the eye of the beholder, so does excitement depend on the excitability of opinion leaders and trend setters. Thus these questions should be asked:
 - 2.1.1. What are the origins of the thinking behind AI?
 - 2.1.2. What are the motivations and intentions, besides the necessity to make money?
 - 2.1.3. Who benefits from the automation? On what levels?
 - 2.1.4. Are the benefits mainly economical?

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?
 - 3.1. Advertising prepares the general public for consumption. If the Committee feels able to counteract this trend, it would be highly desirable!
4. Who in society is gaining the most from the development and use of artificial intelligence and data?
 - 4.1. Big corporations, e.g. Oracle who buy 40 companies a week.
 - 4.2. Who is gaining the least?
 - 4.2.1. Consumers.
 - 4.3. How can potential disparities be mitigated?
 - 4.3.1. By re-dressing imbalances of the money supply into the economy. See <http://www.forumforstablecurrencies.org.uk/>

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?
 - 5.1. The public is not a one-dimensional market addressed by mainstream media, but includes also internet users who have learned to educate themselves.
 - 5.2. Software and AI developers belong to a new kind of 'breed' in the public who, by controlling computers, are far more conscious of society's financial and control mechanisms.
 - 5.3. With a view to the Committee's recommendations, it could maybe establish test centres with exhibitions of ethical AI products to discourage negative uses. The exhibition about robots at the V & A is one small example.
 - 5.4. The various Catapults could be used for that purpose.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?
 - 6.1. AI and ML [machine learning] are typical for the human mind that creates 'because it can', not necessarily because it is useful, ethical, practical, helpful or desirable.
7. How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed?
 - 7.1. How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?
 - 7.2. Open Data is a laudable initiative, but not yet strong enough to counteract any negative implications of AI and ML.
 - 7.3. If, however, it would be presented as a shining example, the Government could become 'digitally credible', after having lagged behind in the digital arena.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence?

- 8.1. How can any negative implications be resolved?
 - 8.2. All technologies are tools that can be used for good and for evil. First, weapons were invented, then financial wars broke out. Now wars take place on mental and psychological levels.
 - 8.3. I am not only following the appeal of the British Computer Society 'to make IT good for society', I am also 'doing good with IT'.
 - 8.4. Furthermore, the Specialist Law Group is concentrating on Electronic Law and Digital Evidence.
9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable?
- 9.1. When should it not be permissible?
 - 9.2. Blackboxing only shields code, not its effects. Either the software is allowed to run and its output be sold or not.
 - 9.3. It is hardly conceivable to establish test centres that check all AI. But it is possible and necessary to watch Big Brothers such as Google, Facebook and Twitter – not only with respect to their taxation!

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom?
- 10.1. Should artificial intelligence be regulated?
 - 10.2. If so, how?
 - 10.3. Society's challenge has shifted from regulation and law to governance and ethics. Electronic laws and digital evidence are becoming more relevant and effective than regulation has ever been.
 - 10.4. As long as money is used as a carrot and a stick, its self-perpetuating destructive usage will be more damaging than AI that is only an outcome of a culture that has been created by greed and false capitalist values.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

- 11.1. An IT officer of the Police told me at a techUK meeting that they know they cannot live without the creativity of SMEs.
- 11.2. Can lessons be learned from creative SMEs rather than institutions that perpetuate existing thinking and cultures? A Google search will always result in more and possibly relevant answers than any consultation.
- 11.3. Chance and coincidences play a big role, when searching, but 'hidden AI' influences such Google activities.
- 11.4. A small positive example is www.sciencedisrupt.com besides my own work which I'd gladly present in oral evidence – with my very best wishes for the consultation.

7 August 2017

Professor Andrew McStay – Written evidence (AIC0015)

“AND THEN THERE’S EMOTIONAL AI...”

AUTHOR

Andrew McStay is Professor of Digital Life, Bangor University, Wales, UK.

BACKGROUND

McStay’s expertise is in the social implications of digital media technologies, privacy and commercial uses of personal data. His recent work has focused on what he terms ‘emotional AI’ and ‘empathic media’. This has involved interviewing over 100 companies, organisations, policy actors and others interested in how technologies interact with human emotional life. He has also conducted survey work to assess how UK citizens feel about technologies that employ data about the body to infer how feel.

KEY RECOMMENDATIONS

1. Committee members should reflect on the social desirability of ‘machine-readable’ emotional life.
2. While there is certainly scope to connect information about emotions with personal data, urgent attention should be paid to practices that passively read expressions and emotional behaviour.

SUBMISSION CONTENTS

1. Introduction
2. The technologies
3. Contemporary context
4. Why is it increasingly being used?
5. Implications, opportunities and risks
6. Policies and rules
7. Recommendations

1. INTRODUCTION

This submission understands artificial intelligence (AI) to be *the study of agents that receive percepts from the environment and perform actions*⁸⁸⁷.

1.1 It is notable that throughout the history of AI, the overwhelming emphasis has been on thought and reason. This submission suggests that artificial *emotional* intelligence (emotional AI) is highly under-represented in discussion of the impact of AI.

These systems, that are able to receive precepts about human emotions and perform actions, introduce new social questions. Specifically: it is entirely desirable that machines are able to use sense and 'feel-into' human emotional life? What are the consequences of being able to see, read, listen, feel, classify, learn and interact with emotional life?

1.2. This involves machines reading words and images; and seeing and sensing facial expressions, gaze direction, gestures and voice. It also encompasses machines feeling our heart rate, body temperature, respiration and electrical properties of skin, among other bodily behaviours. Together, *bodies and emotions have become machine-readable*.

Each of these is underpinned by an interest in affective computing, cognitive computing and other approaches that seek to make human-machine interaction more natural. This a field pioneered by Rosalind Picard⁸⁸⁸. However, affective computing is *an example of 'weak AI'* (rather than strong AI). That is, it reads and reacts to emotions, but it does not think and feel itself (arguably unlike the machines in movies such as Alex Garland's (2014) *Ex Machina*).

⁸⁸⁷ Russell, S. and Norvig, P. (2010) *Artificial Intelligence: A Modern Approach* (3rd ed). Englewood Cliffs, NJ: Prentice Hall. p. viii

⁸⁸⁸ See for example Picard's 1997 ground-breaking book *Affective Computing*, or a 2007 book chapter of Picard's titled *Toward Machines with Emotional Intelligence* at <http://affect.media.mit.edu/pdfs/07.picard-EI-chapter.pdf>.

2. THE TECHNOLOGIES

In as far as AI systems interact with people, one might reason that AI agents have no value until they are sensitive to feelings, emotions and intentions. This includes home assistants and headline grabbing humanoid robots, but the important development is how *emotion recognition systems are progressively permeating human-computer interactions.*

2.1 The following in the list are not in themselves intelligent agents, but they allow intelligent agents to discern and sense people's emotions.

- Online sentiment analysis: this analyses online language, emojis, images and video for evidence of moods, feelings and emotions (typically applied to social media).
- Facial coding of expressions: this analyses faces from a camera feed, a recorded video file, a video frame stream or a photo.
- Voice analytics: the emphasis is less on analysing natural spoken language (what people say), but how they say it. These include elements such as the rate of speech, increases and decreases in pauses, and tone.
- Eye-tracking: this measures point of gaze, eye position and eye movement.
- Wearable devices sense:
 - *Galvanic Skin Response (GSR)*: sweating on hands and feet is triggered by emotional stimulation. Due to the change of balance in positive and negative ions in the sweat, electrical current flows quantifiably differently.
 - *Electromyography (EMG)*: this measures muscle activity and muscle tension, which has been shown to correlate with negative emotions.
 - *Blood Volume Pulse (BVP)*: this bounces infrared light against a skin surface to measure the amount of reflected light to assess heart rate.
 - *Skin temperature (ST)*: sensors measure skin temperature because change in emotion is associated with an increase and decrease in temperature.
 - *Electrocardiogram (ECG)*: sensors measure the heart's beats and rhythms by means of its electrical conduction. Changes in rhythm are associated with change in emotional states.
 - *Respiration rate (R)*: spirometers measure inhalation and exhalation. Changes in respiratory behaviour correlates with emotional states.
- Electroencephalography (EEG): in-house and wearable approaches record electrical signals along the scalp to measure brain activity.
- Gesture and behaviour: cameras are used to track hands, face and other parts of the body said to communicate particular messages and emotional reactions.
- Virtual Reality (VR): this an experience in which people cede themselves to varying degrees to a synthetic environment. It allows remote viewers to understand and feel-into what the wearer is experiencing.

- Augmented Reality (AR): this is where reality is overlaid with additional computer-generated input (such as graphics). Remote viewers can track attention and interaction with these digital objects.

2.1 On whether machines can really understand emotions, we have two possibilities: 1) *genuine understanding*; and 2) *simulated understanding* (the capacity to approximate, to contextualise within what one can comprehend, to make educated judgements and to respond in an appropriate manner). Much of this has been rehearsed in debates about whether machines can really think, what it is for a person to think and the philosophical knots associated with knowing the lives of others. Possibility 2 is a reasonable proposition: if we allow for the possibility of a simulated and observational version of emotional life, this provides scope for machines to interact with people in meaningful ways.

3. CONTEMPORARY CONTEXT

Small and medium-sized enterprises (such as Affectiva in the US and Sensus in the UK) have adopted affective technologies and applied them for diverse ends to gain first-mover advantage, but in recent years large technology companies including Apple⁸⁸⁹, Facebook⁸⁹⁰, Google⁸⁹¹, IBM⁸⁹² and Microsoft⁸⁹³ are becoming more public and active in their emotion and cognitive applications. This will bring about a net rise of emotion-enabled interaction. Notably, the 2016 Gartner report on influential emerging technologies pitched affective computing as being between 5-10 years away from mainstream adoption (Gartner, 2016). This aligns with my findings from interviewees, who are all confident that emotion detection would increase in prevalence, breadth of applications and capacity to engage with the wider context of our lives in 5-10 years.

3.1 Who uses it?

Emotional AI has scope to be employed in any context where it may be useful to know how a person or group of people is feeling. This encompasses self-tracking (such as with wearables to gauge one's own mental or physical health) as well as organizational monitoring. A non-exhaustive range of sectors and organisation types interested in acquiring insights into emotional life is outlined in Table 1 below.

⁸⁸⁹ In 2016 Apple purchased facial coding company Emotient and, separately, has logged patents for mood-sensitive television advertising. Patent available at: <http://tinyurl.com/zm64n2s>

⁸⁹⁰ Facebook famously experimented with mood contagion through their social networks in 2014. In 2017 they reportedly offered advertisers opportunity to target younger users undergoing psychological vulnerability, such as when they felt "worthless," "insecure," "stressed," "defeated," "anxious," and a "failure." See: <https://theconversation.com/tech-firms-want-to-detect-your-emotions-and-expressions-but-people-dont-like-it-80153>

⁸⁹¹ Google <https://cloud.google.com/vision/>

⁸⁹² IBM pursues emotion detection through Watson. I interviewed Watson at SXSW in 2016; Watson also attended my Digital Catapult workshop.

⁸⁹³ Microsoft Azure use emotion products in their computer vision services: <https://azure.microsoft.com/en-gb/services/cognitive-services/emotion/>

Table 1 Sectors and Organisations interested in Emotional AI

Sector/ Organisation	Form of tracking	Reason for interest in tracking emotions
Advertisers	Sentiment, voice, facial coding and biometrics	To understand preferences and behaviour; and to optimise creative components of adverts and targeting.
AI/cognitive services	Sentiment, voice, facial coding and biometrics	Wide-ranging, but in general to enhance interaction with devices, services and content.
Artists	Sentiment, facial coding and biometrics	To create artwork, measure engagement and to expand their creative toolbox.
Automobiles	Facial coding and biometrics	To assess behaviour and stress, using this to inform insurance decisions.
Brand marketers	Sentiment, facial coding and biometrics	To understand social discussion and reactions to brand content (logos, products, messages, etc.).
City experience analysts	Sentiment, facial coding and biometrics	To gauge citizen feeling about initiatives.
Data brokers ⁸⁹⁴	Sentiment, facial coding and biometrics	Early days for this, but scope to understand how consumers feel is valuable.
Education	Facial coding and biometrics	To analyse in-class behaviour and individual comprehension and engagement with content.
Financial trends	Sentiment	To chart market emotionality (via assessment of social media influencers, discussion and media coverage).
Gaming	Facial coding and biometrics	Used as input devices to enhance gameplay.
Health	Sentiment, voice, facial	A social media profile of a person may be tracked for indication of mental state,

⁸⁹⁴ They aggregate, process, clean, analyse and license information. This includes marketing and consumer trends data, but it also encompasses commercial data about companies, scientific and technical data, real estate information, or geo-location data.

	coding and biometrics	and wearables provide biometric insight into short and long-term states.
Home Internet of Things (IoT)	Sentiment, voice, facial coding and biometrics	Patents by large technology companies show interest in domestic emotion tracking to personalise services and advertising. Examples include home voice assistants, voice-activated devices and cameras on top of televisions).
Insurance	Sentiment, facial coding and biometrics	To help understand customer emotional disposition and mental health (risk assessment).
Police forces	Sentiment	To gauge civic feeling and identify disturbances before they occur.
Police/security technologists	Sentiment and biometrics	To assess citizens (via sentiment analysis) but also officers (via biometrics, e.g. stress trackers).
Political parties/organisations	Sentiment	To gauge reactions to policies and government/party initiatives.
Product/user testing	Sentiment, facial coding and biometrics	To assess reactions to products and specific features.
Robotics	Facial coding and voice	To enhance interaction between robots and people.
Sextech	Biometrics	To enhance sex life, make devices/robots more responsive.
Social media companies	Sentiment and facial coding	To assess sentiment, emoji usage, group behaviour, individual profiling, altering and posting behaviour. Companies are patenting and experimenting with facial coding, heart rate sensing and voice analytics.
Television/movie content developers	Sentiment, facial coding and biometrics	To test reactions to shows/movies.
Retailers	Sentiment, voice, facial coding and biometrics	To assess in-store behaviour (and potential to link reactions with online/loyalty profiles).

Mood tracking wearable developers	Biometrics	So a person or organisation can track the reactions, emotions and moods of a person.
Workplaces	Biometrics	To organisationally track emotions and moods.

4. WHY IS IT INCREASINGLY BEING USED?

Uses include making technologies easier to use, evolving services, creating new forms of entertainment, giving pleasure, finding novel modes of expression, enhancing communication, cultivating health, enabling education, improving policing, heightening surveillance, managing workplaces, understanding experience and influencing people.

4.1 Significance of Emotional AI

Life with 'technologies that feel' brings all sorts of implications, but they all spring from the fact that *people and emotional life are increasingly machine-readable*. This is a considerable development from the online tracking that most people are familiar, if not comfortable, with. The opportunities and threats are two-fold:

4.2 On the one hand, it *naturalizes interaction with technology*. This has scope to enhance interaction with our personal devices, make them more responsive to our wants and needs, provide novel forms of entertainment, increase enjoyment of existing content and media, and positively assist with education and health.

4.3 On the other, it provides scope for *surveillance and exploitation of emotional life*. The concern is arguably less about the act of watching itself, but rather: a) privacy concerns; b) whether citizens are happy about passive tracking of their emotions; and c) questions regarding how this information is used. The scope for ubiquitous monitoring should be understood in terms of apps and personal devices, domestic spaces, workplaces and quasi-public spaces (such as retail outlets). I do not use words such as 'surveillance' or 'exploitation' in an alarmist fashion, because it seems entirely sensible to suggest that emotion capture technologies will be used in any context where it is useful to understand how a person or group of people feel. This has special application to the advertising, marketing and retail industries (and the media and technology platforms that serve them) because all have a long interest in psychological profiling of citizens.

5. IMPLICATIONS, OPPORTUNITIES AND RISKS

5.1 Opportunities

Although this submission is attentive to data protection and privacy harms, the scope for positive benefits should not be missed. With appropriate data protection, there is significant scope to use emotional AI to create more fulfilling media experiences. Gaming, for example, is enhanced with eye-tracking, biometric input and game engines that learn more about their players. There is scope for use in health (personal self-care through wearables), but also institutional healthcare (physical and mental). Architecture and city planning may benefit from understanding reactivity to generate emotional topologies and maps. There is significant scope for artistic exploration of place, objects and events (potentially with citizen/co-created annotation). Concern should not focus on the technologies per se, but their deployment.

5.1.1 Positive usage requires *responsible innovation*. As outlined below in Section 5.3, people are increasingly open to new modes of biometric interactivity, but they are wary. A responsible approach requires consideration of value-sensitive design, privacy-by-design and applied ethics. This recognises that technologies are not socially neutral and that the nature of how technologies are designed, rolled out and employed has social consequences. A responsible approach to innovation embeds positive social values into technology.

5.1.2. Responsible innovation will not happen by simply asking innovators to be responsible. It requires advice and pressure from regulators, data protection bodies, research funders, start-up incubators, industry and corporate leaders, smart city vendors, municipal managers, NGOs, and universities through both teaching and research. Regulators and data protection authorities, for example, might interact at early stages, offer advice and guide innovators on what their stance is likely to be. Research funders can incentivize scientists and innovators to insist that data ethics are meaningfully built into funded design processes, and universities can insist that technology courses contain ethical considerations. Similarly, incubators and corporate leaders (from regional innovators to large bodies such as the World Economic Forum) may advise on thinking through potential social consequences of technological development. This is certainly not a silver bullet solution, but by initiating conversation about the implications of mining highly personal and intimate data, we improve our chances of these technologies being individually and socially beneficial.

5.1.3 An ethically led approach to this sector may be the one that succeeds in the marketplace. Younger citizens are not against the principle of emotion tracking and new modes of interactivity, but they appear to be rightfully wary⁸⁹⁵. Rather than relying on citizen habituation to uncomfortable and unwanted conditions, a responsible approach is one that promotes creativity, fun, reward,

⁸⁹⁵ Survey details (n=2067) available from McStay, A. (2018) *Empathic Media: The Rise of Emotion AI*. See: <https://drive.google.com/file/d/0BzU2NrGCFp7qd0tOWjJDcFgxdGc/view>

benefits and a very upfront approach to why data is being used and what happens with it.

5.2 Risks

Opportunities and benefits may quickly become threats, especially if people feel out of control, or undergo learned helplessness in face of technology and data collection/processing about emotional life. Any interest in making emotional life and the body machine-readable should be treated with the highest levels of caution. I suggest that there is something fundamentally important about emotional life and its centrality to human experience. Any scope to commoditise emotions must be treated critically and carefully. The task is to find appropriate means of living with emotion-capture technologies in a way that respects the dignity of human life, enhances experience of technologies, and serves rather than exploits people. Risks include the following:

- People are treated as emotional animals to be biologically mapped and manipulated.
- People are seen as objects rather than as subjects.
- People do not have control over sensitive information.
- Passive tracking collects intimate data without consent.
- Alienation from public spaces.
- Unwanted attention to behaviour and the body.
- Increased scope to manipulate consumer behaviour.
- Workplace coercion.
- Inadequate data protection coverage (this tends to focus on questions of identification rather than matters of dignity and the body). More on this in Section 6.

5.3 Do people care?

In 2015 I ran a UK survey (n=2067) to gauge attitudes towards the potential for emotion detection in a range of then nascent everyday uses of emotional AI⁸⁹⁶. These were sentiment analysis, out-of-home advertising, gaming, interactive movies, and voice-based capture through mobile phones.

5.3.1. Overall findings were not significantly different between each of the proposed emotion detection methods. The overall figures derived from reactions to each method are:

- 50.6% of UK citizens are 'not OK' with emotion detection in any form;
- 30.6% are 'OK' with emotion detection if they are not personally identifiable;
- 8.2% are 'OK' with having data about emotions connected with personally

⁸⁹⁶ For full results see McStay, A. (2018) *Empathic Media: The Rise of Emotion AI*. Available from: https://www.researchgate.net/publication/317616480_EMPATHIC_MEDIA_THE_RISE_OF_EMOTION_AI

identifiable information;

- 10.4% do not know.

A key outcome is that the significance is that *only 38.8% of UK citizenry surveyed can be said to be 'OK' with having any data about emotions collected about them.*

5.3.2. Also of interest is age. While gender, social class and region did not produce noticeable variances from their respective mean averages, age did produce significant deviations. Younger people (18-24) were more likely than any other age group to be 'OK' with some form of emotion detection in the digital media and services they use. To illustrate, the mean average of all age groups 'not OK' with *any* form of emotion detection is 50.6%. However, for 18-24s this is only 31.2%. In contrast, over 65s are least likely to be 'OK' with it: 62.2% are 'not OK' with it.

5.3.3 However, there is another key insight: few of any generation are keen on having data about emotions linked with personally identifiable information. The averages for this are low beginning at 13.8% for 18-24s who are 'OK' with it, and descending to 1.6% for people 65+.

5.3.3.1 I reason that young people are willing to accept higher levels of profiling, but companies and other organisations should *not* read this as, 'young people don't care about privacy'. I suggest *younger people are open to new media experiences, but they seek control over the process.*

6. POLICIES AND RULES

6.1. In Europe, 'personal data' is that which identifies a person or singles them out as a person for unique treatment, whereas 'sensitive data' includes information about the body, ethnicity, political opinion, trade union membership, sex life and offences (past or pending). As such, biometric data, when 'personal', is considered as sensitive.

6.2. Importantly however, *biometric information about emotion that does not identify or single-out a person does not have legal coverage.*

6.3. Legal experts and data protection policy-makers I met from the European Commission agreed with this assessment and that this is an unanticipated lacuna. The consequence is that anonymous emotion tracking may take place without consent by citizens.

6.4 However, there is potentially scope within the EU General Data Protection Regulation (see Articles 6(2) and 9(4)) that allows Member States to introduce further conditions regarding biometric data. I suggest that greater civic conversation is required about the desirability of making emotional life machine-readable. Put otherwise, *should we have privacy considerations based on intimacy rather than solely about whether a person is identifiable or not?* Are there questions to be asked regarding respect for human dignity?

6.5 Case examples might include cameras that scan data points on a person's face in a retail outlet to discern emotional behaviour (such as eye, lip and nose movement). The same applies to out-of-home digital advertising where cameras above ads at shelter ads have scanned people for emotional reactions⁸⁹⁷. There is also considerable in using wearable devices at work to gauge emotional behaviour at work (such as to gauge stress and performance of workers). While that which identifies people is subject to high data protection provision, aggregated and non-identifying data is not⁸⁹⁸.

6.6 Self-regulators also have a role to play in that, on the basis of the survey work conducted for this study, citizens are wary of emotion tracking practices. Self-regulators (especially in advertising) should consider extending protection of citizens beyond existing data protection laws (that focus on identification) to encompass intimacy, dignity and the desirability of making emotions machine-readable in public places. This is especially the case given their remit of social responsibility.

⁸⁹⁷ McStay, A. (2016) Empathic media and advertising: Industry, policy, legal and citizen perspectives (the case for intimacy), *Big Data & Society*, (pre-publication): 1-11. Link: <http://bds.sagepub.com/content/3/2/2053951716666868.full.pdf>

⁸⁹⁸ McStay, A. (2017) Wearables-at-Work: Quantifying the Emotional Self, *Privacy Laws and Business*. Link: https://www.researchgate.net/publication/317490573_Wearables-at-work_quantifying_the_emotional_self

7. RECOMMENDATIONS

The most immediate concern is the need to tackle the fact that *legal consent is not required to capture data about emotions that is not personal (i.e. capable of identifying or singling-out a person)*.

7.1 This is a lacuna which will be exploited, perhaps especially by retailers and advertisers using computer vision techniques in public and quasi-public spaces. Data protection authorities and industry self-regulators across diverse sectors (such as advertising, consumer protection, retail and marketing) need to tackle the following question: *Beyond the law as it stands today, are citizens and the reputation of the industries that self-regulators are charged to protect, best served by covert surveillance of emotional life?*

7.2 If the answer is no, they should immediately amend their codes of practice. The reason is that questions of ethics, emotion capture and making bodies passively machine-readable by emotional AI is not contingent upon personal identification, but human dignity, choice and decisions about what kinds of environments we want to live in.

7.3. Difficulties with regulation are noted, but nonetheless meaningful legal and self-regulatory scrutiny is required. To mitigate potential for poor regulation, industry along with other stakeholders (policy, regulators, data protection NGOs and civil society) are encouraged to openly discuss, debate and take a leading role in shaping emotional AI and affective technologies in such a way that these technologies can be agreed by a wider set of stakeholders to serve rather than exploit people.

17 August 2017

medConfidential – Written evidence (AIC0063)

medConfidential submission to the House of Lords Select Committee on Artificial Intelligence

Cover sheet

Our submission is structured with an overview, and then answer the questions from the Committee. In part 2 we provide context for our answers in light of the deal between the Royal Free Hospital Trust in London and Google DeepMind that was found to be unlawful – resulting from a complaint lodged by medConfidential. Part 3 looks to wider lessons for the public and private sectors and the culture of data and digital services in the current environment.

About medConfidential

medConfidential is an independent non-partisan organisation campaigning for confidentiality and consent in health and social care, which seeks to ensure that every flow of data into, across and out of the NHS and care system is *consensual, safe, and transparent*.

Founded in January 2013, medConfidential works with patients and medics, service users and care professionals; draws advice from a network of experts in the fields of health informatics, computer security, law/ethics and privacy; and believes there need be no conflict between good research, good ethics and good medical care.

medConfidential's core work is funded by an annual grant from the Joseph Rowntree Reform Trust Ltd, which for 2017-2018, is under £50,000 and covers two staff. We also accept donations in aid of our wider work across data,⁸⁹⁹ of which AI forms one small part.

Evidence from medConfidential: Overview

- 1 Healthy, successful, and particularly male protagonists commonly have no understanding or insight into the types of sensitive information they will one day have to divulge to their doctor, and the consequences of confidentiality not being respected.
- 2 In any AI future, doctors remain doctors. Patients retain the fundamental right to make decisions about their care – a patient has the right to make an informed decision to reject life-saving care. It is perverse that some would argue patients get less control over their data than they do over their treatments – no one advocating that will have passed medical school.

⁸⁹⁹ <https://medConfidential.org/donations>

- 3 To quote a former Director of GCHQ: “We have learnt from the tech sector that expertise needs to be at the heart of strategy. Relying solely on the well-meaning generalist, which has not served government policy well in computer science since the 1950s, is not enough.”⁹⁰⁰ Issues are raised, not because AI models are so hard to create that only the current leaders can do it, but as these models are so easy to train when you have domain-specific knowledge that AI will become a commodity. We are not treating it as such, or considering that possibility. One reason large players would wish others to believe AI training is hard is because, generally lacking specialised knowledge of a domain, it is hard for *them*. Not everyone chooses those disadvantages – an NHS clinician with technical skills and some AI training would probably be far better equipped to build tools for anything within their speciality.
- 4 AI is not magic. AI bestows on its creators, users, and victims no capability that is not data processing. It may be novel data processing, it may be highly processing-intensive data processing, but it remains just data processing. We have laws for that. Our laws should be adequate, where both the letter and the spirit of them are followed.
- 5 Data controllers remain data controllers. Today, and more so under GDPR, data controllers should tell their data subjects how the data they hold has been used. Dodgy decisions are far less likely to get made when the decision-maker knows that every person whose data is affected will be told that it was used in such a way.
- 6 Principles endure. Our human rights laws, and our data laws, are no more in need of ‘update’ for AI than they were for any other new technology. Some, with self-interested motives may wish to undermine laws and human rights. That they desire greater private benefit, at the expense of the public interest, does not mean that they should succeed.
- 7 medConfidential works for systematic improvements, which require constructive proposals. We have previously published, after discussions with a number of AI companies, thoughts around contributing to resilient public trust via transparency,⁹⁰¹ and making audited changes in an accountable way.⁹⁰² AI will change society, and society should account for that.
- 8 For decades the NHS has understood that sharing data raises ethics and consent issues, and sharing AI models will become no more technically complicated than sharing data is today. While there are a number of

⁹⁰⁰ <https://www.ft.com/content/92a651d0-05bb-11e7-aa5b-6bb07f5c8e12>

⁹⁰¹ <https://medconfidential.org/wp-content/uploads/2017/05/1-resilient-public-trust.pdf>

⁹⁰² <https://medconfidential.org/wp-content/uploads/2017/05/2-modifying-audit.pdf>

notable examples of data sharing not always being done properly in practice, there is generally good intent across the NHS as a whole – underpinned by a belief that patients should be informed of what their choices are. Since the care.data project collapsed in 2014, the NHS, led by the Secretary of State, has been putting in place the infrastructure necessary to support informed choices, and those structures will apply to AI uses. The rest of Government has barely begun; the private sector mostly doesn't yet believe it needs to.

One action for the NHS

Q10:

- 9 The most life-changing, rapid, and one-off decisions people must make are those to do with their health, and the health of their loved ones. In such situations, the benefits of diversity – of clinicians and patients seeing complex issues from different perspectives – are well understood.
- 10 In medicine, there is a culture of “second opinions” – as a patient, you can always ask another doctor for their independent opinion on an issue. This is acknowledged as a great strength of the medical community; indeed, the seeking of diverse (even possibly contradictory) opinions is actively supported by professionals realistic and humble enough to accept that there may not be one single right answer.
- 11 As technology progresses, why would we choose a *lower* standard for AIs offering diagnostic assistance to doctors?
- 12 To ensure this diversity survives, **for each clinical speciality, the NHS and research bodies should support two PhDs / postdocs within the NHS to build AI assistance tools that will help the NHS front-line, and additionally, mandate that any AI clinical assistance funded by the NHS must be the composite of 3 different AI models, trained on different data sets.**⁹⁰³
- 13 Producing this within the NHS means existing research processes can be used, and clinical standards will be met. The mandate that there must be at least 3 independent tools assisting in any diagnosis also provides a clear demonstration to the private sector that the NHS is committed to avoiding a monopoly on AI suppliers, including its own.
- 14 AI appears now to be as fundamental to future health diagnosis as DNA was seen in the 90s – not just in the UK, but worldwide. The Human Genome Project was committed to public service – when a opportunistic venture capital-backed private company attempted to replace the project

⁹⁰³ <https://medconfidential.org/2017/everyones-experience-in-ai-decision-making/>

with its own private monopoly,⁹⁰⁴ scientists and charities committed to the viability of freely available data and tools for the public interest. With the rush to commercialise the decision-making for basic diagnoses, we are facing a similar threat.

Part 1 Committee questions

Q2, 7:

15 We take it as given that there are useful applications of AI, both in health and beyond. The question of AI as it relates to medConfidential's work is whether such uses will be done in a manner which is consensual, safe, and transparent. Just as with any other new technology, the question is how we choose to use it; we should choose wisely. We will eventually do the right thing – what is uncertain is how many attempts it will take.

16 Concerns about Amazon, Google, Facebook, *et al.* gaining a monopoly over AI are based on them being the data controller for a monopoly that has an economy of scale, not only as a data processor. Controversy of a monopoly which each hopes no one else can have, which can only exist if it is impossible to replicate the tools, and copying can be restricted. Part 2 of this submission is a case study where replication is easy, and where one large AI provider states this to be true.

17 Commodity services can be outsourced, a trade secret or unique competitive advantage won't be. While in the private sector, data controllers and data processors will often be the same entity, in the public sector, they probably will not be. Data controllers retain the right to use a different and competing data processor should they wish to. And in that difference comes the public sector's ability to repeat the training of an AI model with a new provider, should it wish to.

Q4, 6, 7, 9:

18 AI development is evolving at a rapid pace, but as a commodity, there is no possibility of the public sector being "left behind" because it can simply replicate/use anything that works. It retains the data controllership and can 'catch up' by picking a supplier, after something has been shown to work elsewhere. Training AI models is cheap and easy,⁹⁰⁵ once you know

⁹⁰⁴ At the time, the BBC talked to both projects: <http://news.bbc.co.uk/1/hi/sci/tech/716479.stm> and <http://news.bbc.co.uk/1/hi/sci/tech/716613.stm>. This paper also discusses the public/private funding issues: "Realities of data sharing using the genome wars as case study - an historical perspective and commentary" <https://epjdatascience.springeropen.com/articles/10.1140/epjds13> <https://doi.org/10.1140/epjds13>.

⁹⁰⁵ This is also not, entirely, positive: <http://www.disruptiveproactivity.com/2017/08/ai-in-the-school-playground/>

how to train the model (which is really hard), and that it *can* be done (as Google DeepMind showed with AlphaGo).

- 19 The capability that is most jealously guarded amongst commercial entities is their proprietary training datasets – that is what makes Google, Facebook, and Amazon unique. If a training dataset is simultaneously unique, impossible to duplicate, and held on commercial terms, it may be felt to be an unassailable lead.⁹⁰⁶ In the public sector, these concerns simply do not arise, unless created by choices in contracts (often with private companies who seek to impose their business model or culture on public sector customers).
- 20 The problems of choosing AI lie not in the tool itself, but in the human decisions about governance and priorities, which may be ill-considered or perverse. When considering inclusion within training datasets, it is important to ask questions about who could be in the training dataset. **All is different to most; most is different to some; and some is very different to all.**
- 21 Approaches that demand different volumes of data should be regulated differently. Algorithms requiring “all” data should be fully transparent and intensively regulated; “some” data can be lightly so, unless they have significant human impact (e.g. data about people). We already treat a census differently to a mostly-representative survey, differently to an opinion poll – there are well-founded statistical reasons for doing so, and those same reasons apply to AI. Many of the poor business decisions in the use of AI come from a misunderstanding of these different categories. “All” mandates every edge case and law must be considered⁹⁰⁷; “most” allows people a choice; “some” may entirely ignore minority groups.⁹⁰⁸
- Q2, 3, 5:
- 22 The AI industry’s catch-all term for concerns such as these is “AI safety”. It is unclear what will go wrong, how and when, and what public concerns will be as a result. It is only certain that, in the long term, something *will* go badly.⁹⁰⁹ What happens afterwards?
- 23 In some quarters, there is an absolute belief in the supremacy of technical systems. An assumption by advocates that the decisions of human

⁹⁰⁶ Humans also felt humans had an unassailable lead in Go; until it was clear we didn’t have a lead at all.

⁹⁰⁷ Including, and especially, Human Rights.

⁹⁰⁸ Which leads to: <http://www.mirror.co.uk/news/world-news/racist-soap-dispenser-refuses-help-11004385>

⁹⁰⁹ There is a tendency in some businesses to hope that it happens to a competitor first; or to rely on the hope that the problem will be one for their successor or liquidators.

creators are incapable of error or omission, that technological determinism will ensure that nothing could ever go wrong or stray beyond expectations, and that there will be no unpredictable human response⁹¹⁰ to or critique of their actions.⁹¹¹ Over time, reality will intervene and show this belief to be entirely false. A company telling people that some very complicated cryptography confirms they are 'telling the truth' will not satisfy an unhappy or fearful public.

- 24 Human creators of tools may not have considered how their creations come up with the outputs they do. After all, producing an enthusiastic press release that makes a company sound good is far easier than finding out how it could have been replicated by anyone – and would also undermine claims of magic and perfection.
- 25 We do not mandate that pharmaceutical drugs must be perfectly safe – we have a yellow card scheme⁹¹² for reporting adverse events, and the system learns from them. The same mechanism should apply to rare exceptions in AI tools, as the system sometimes being wrong is a natural output of any imperfect system.
- 26 For the people who are making decisions about the future of artificial intelligence, a lack of humility and introspection and, above all, an inability to proactively and honestly admit error, is of deep concern. The most disturbing part of the whole RFH/DeepMind fiasco is not that it happened, but that it was unable to acknowledge an error without being forced to do so by regulators.
- 27 The biggest risk medConfidential sees from AI is not hostile entities using such tools with aggressive intent;⁹¹³ it is that good people will do what they feel is a good thing, with initially trivial negative consequences – followed by the very human instinct to cover it up.

Part 2: Practical experience of an AI company's use of 1.6 million patients' NHS medical records

⁹¹⁰ "Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms" <https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms>

⁹¹¹ A hypothesis instantly disproved by anyone who has understood '2001: A Space Odyssey', or many other films.

⁹¹² <https://yellowcard.mhra.gov.uk>

⁹¹³ We have confidence in the Civil Contingencies Secretariat and the Ministry of Defence; although whether anything of our house will remain when they get there is unclear: <https://www.dailydot.com/debug/google-deepmind-ai-learns-to-walk/>

28 It is a matter of public record that Google DeepMind and the Royal Free Hospital had an unlawful agreement in this case.⁹¹⁴ We do not propose to repeat the detailed evidence that we provided to previous investigations here, but it is available in various documents:
medConfidential letter to Regulators - June 2016⁹¹⁵
Timeline of events (up to June 2016)⁹¹⁶
Letter from National Data Guardian to DeepMind/RFH - February 2017⁹¹⁷
medConfidential update after publication of the NDG letter - May 2017⁹¹⁸

29 In general, the culture clash between the NHS and the technology industry could be described thus: the technology industry often uses well-meaning and very well paid disciplinary generalists, for which health is just today's task; whereas the NHS approach has traditionally relied upon underpaid disciplinary specialists who care about the work they do. Both are choices, though they may not interact entirely well.

2.1 How did it go wrong?

30 A doctor who wanted to 'be first',⁹¹⁹ in an institution which prioritises being first over being right, cut corners. Google DeepMind, which also has a need to 'be first', didn't understand what the agreement it signed actually required it to do. "Direct care" has a particular legal definition – it appears no one at Google DeepMind thought to put the term into a search engine to check what it meant, and whether there were any rules they should have followed. So they followed none of them.

31 In late 2015, Google DeepMind made a request to the Health Research Authority to do "machine learning" research on the data it has from the Royal Free. The application said:⁹²⁰

"DeepMind acting as a data processor, under existing information sharing agreements with the responsible care organisations (in this case Royal Free Hospitals NHS Trust), and

⁹¹⁴ <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law/>

⁹¹⁵ <https://medconfidential.org/wp-content/uploads/2016/06/medconfidential-to-regulators.pdf>

⁹¹⁶ <https://medconfidential.org/wp-content/uploads/2016/06/medconfidential-deepmind-timeline.pdf>

⁹¹⁷ <http://news.sky.com/story/google-received-16-million-nhs-patients-data-on-an-inappropriate-legal-basis-10879142>

⁹¹⁸ <https://medconfidential.org/wp-content/uploads/2017/09/2017-06-10-Deepmind-NDG.pdf>

⁹¹⁹ <http://www.bbc.co.uk/news/technology-36783521>

⁹²⁰ Page 17

<https://www.whatdotheyknow.com/request/410881/response/1001252/attach/2/1718%20FOI%20011%20HRA%20response%20and%20documentation.pdf>

providing existing services on identifiable patient data, will identify and anonymise the relevant records.”

- 32 Google DeepMind subsequently made public statements that those “existing information sharing agreements” do not allow it to do⁹²¹ precisely what it previously told an ethics board it *could* do under those agreements.⁹²² Which of those statements is true matters less than the inconsistency between supposedly truthful statements.
- 33 When the “DeepMind Health Independent Reviewers” issued a report in June 2017,⁹²³ it made no mention of this request, which started in 2015 and seemingly continued into June 2016. We do not know why, and (at the time of writing) the Reviewers have not answered that question. In an industry where a major player has a motto of “move fast and break things”, it is unclear what happened to that project in the intervening months.
- 34 At the time of writing,⁹²⁴ Google DeepMind refuses to answer the question, “**Did you feed the data to your AI?**”. The company has not always been so reticent. In the early days of the press coverage into this deal, Google was told by the Royal Free that the data it would have fed to the AI was “outside of Caldicott”. While this is not correct, Google would have believed it was when it made statements that there was “no AI”, and that the project was “not research” but only “direct care”. Therefore, based on Google DeepMind’s own statements, medConfidential believes it likely that Google DeepMind did feed 1.6 million patients’ medical records to its AI, because the company was not aware it was unlawful and unethical to do so. We expect its officers thought they did not need to disclose what they had done with the 1.6 million records, when asked in mid-2016 – the Silicon Valley culture of secrecy overriding the public’s right to know how data about them has been used by contractors to public bodies.

2.2 AIs are easily copyable

⁹²¹ <http://www.bbc.co.uk/news/technology-39301901>

⁹²² Page 17

<https://www.whatdotheyknow.com/request/410881/response/1001252/attach/2/1718%20FOI%200011%20HRA%20response%20and%20documentation.pdf>

⁹²³

<https://deepmind.com/documents/85/DeepMind%20Health%20Independent%20Review%20Annual%20Report%202017.pdf>

⁹²⁴ As we write, this piece came out yesterday: <https://techcrunch.com/2017/08/31/documents-detail-deepminds-plan-to-apply-ai-to-nhs-data-in-2015/>

35 When we requested the paperwork given to the ethics board for this project, for supposedly public benefit research, the purposes of the project were redacted with a very unusual justification:⁹²⁵ [emphasis added]

“FOIA Section 43(2) exemption; **commercial interests, has been applied to a number of question detailed in the NHS REC form (iv) and the study protocol (v). Information referenced in the document is commercially sensitive as the release of the information ‘would, or would be likely to, prejudice the commercial interests of any person’.** The relevant information redacted from the application form and detailed in the protocol relates to the **research methodology which, if shared, could potentially lead to the replication by a competitor organisation** which would prejudice the commercial interest of Google DeepMind.

36 That Google demanded methodology be redacted to avoid replication indicates how easy replication would be. The very definition of science, and an underpinning of medicine, is an open methodology that others can copy. We should be careful what we throw away in a rush to commercialisation.

2.3 Business models

37 As stated earlier, medConfidential does not believe there can be a monopoly of private data providers within the public sector, unless extremely poor contracts are signed. While this is almost certainly the case in various instances, it is not a concern unique to AI, and monopolistic and predatory suppliers are not a new phenomenon.

38 After admitting the ease with which their work could be duplicated, DeepMind’s contract requires that the “Streams” app may not be used with any other service. This is ‘lock-in by contract’ – DeepMind’s approach is not so different from Capita.

39 While in publicity, Google DeepMind claims to follow open standards, its contract bans users from connecting to other services. ‘Publicity giveth; the small print taketh away.’ It is in precisely such behaviour that (accidental) monopolies may be created.

40 However, given the openness around the development of artificial intelligence, and the replicability of approaches, any monopoly will persist

⁹²⁵ Page 2.

<https://www.whatdotheyknow.com/request/410881/response/1001252/attach/2/1718%20FOI%200011%20HRA%20response%20and%20documentation.pdf>

only while the institutions which agreed to it remain in the dark (or captured).

41 The same openness that supports long-term public institutions also has other effects. In this Part, we have looked primarily at Google DeepMind. While Google may simply want to use AI to show us better ads on pages we choose to visit, Facebook pokes human psychology,⁹²⁶ for profit (possibly over democracy⁹²⁷), via methods it refuses to talk about,⁹²⁸ after past backlashes against abuses.⁹²⁹ The truth always comes out. Eventually.

42 Our concern is not that an AI will do something dangerous of its own volition, but that – as in the Royal Free and Facebook examples – “AI safety” technical measures alone will be insufficient to deal with human motives or stupidity.

Part 3 - AI in the context of health and the wider public sector

43 Every data flow in the NHS should be consensual, safe, and transparent – whether that data is used in genomics and AI, care.data or the National Data Lake, or whatever comes after genomics and AI.

44 As a result of current and future challenges, every patient should be able to see how their data – data about them – has been used. Such information is necessary for patients to be able to make informed choices as to how their medical records should be used. The information gap between expectations and reality should be closed, and never allowed to open again. This requires ongoing communications and education.⁹³⁰

45 New programmes, whether AI-related or not, are generally coming into established systems and institutions. They do not come in with a blank slate, and without tradeoffs having already implicitly been made.

46 The public sector is a data controller, often statutorily so, and that authority and responsibility cannot be signed away by contract or through commercial desire. Data controllers remain data controllers, which affords

⁹²⁶ <https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/>

⁹²⁷ <https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy>

⁹²⁸ <http://gizmodo.com/facebook-figured-out-my-family-secrets-and-it-wont-tel-1797696163>

⁹²⁹ <https://ethicsandsociety.org/2014/07/01/issues-of-research-ethics-in-the-facebook-mood-manipulation-study-the-importance-of-multiple-perspectives-full-text/>

⁹³⁰ <https://medconfidential.org/wp-content/uploads/2016/07/2015-09-NDG-presentation-shortenedforweb.pdf>

data controllers in the public sector a unique lever that renders moot many of the concerns about “falling behind” on AI. GDPR supports this.

47 In short, no AI use of public bodies’ data will be a monopoly, unless the public body chooses it to be. Public bodies have *service* as their model, not the pursuit of profit.

48 The mechanisms for training an effective AI do not remain secret – this is fundamentally open research.⁹³¹ As such, when companies are able to lock institutions by contract into using only their tools, claims to “use open standards” are entirely misleading, if the tool actively prevents you from using innovations elsewhere. This is a major point of concern with the Google DeepMind app ‘Streams’; not the app itself, but terms of the contract required to use it. The Streams app is easily replicable should any app developer wish to try – the only thing they couldn’t replicate is Google’s PR and lobbying budgets, and halo of influence.

49 PHE and the NHS have found widespread market failure in “apps”. For example, while there are many, many apps to help a woman get pregnant and calculate a due date, there are far fewer high quality support tools covering the period to birth. Market forces, driven by advertiser interest, do not deliver the benefits that are in the wider public interest.

3.1: Why AI companies presume the necessity of ubiquitous surveillance

50 One reason AI companies use games to research and evaluate new approaches is because the scoring mechanism gives an instant, simple, and clear numeric metric of success and improvement. Games do not generally contain real ethical quandaries, and rarely have any real world impacts.⁹³² While they may simulate the real world, they are not the real world; games are used for the same reason we train pilots in simulators.

51 The NHS, and public services, operate entirely within the real world, serving real people. Their lives are not a game, and data from those lives should not be treated with similar disregard. Capturing “all” data is easy in a simulator or game, but not in the real world. In requiring all data and ubiquitous surveillance, we are giving up a great deal for questionable incremental benefit.

52 Unreadable and unread “terms and conditions” allow companies to argue they can do anything. We hold public bodies to a higher standard, and should continue to do so. The NHS and public services operate on a basis

⁹³¹ One of the leaders in the field - OpenAI - was founded on the premise that it would open everything it does.

⁹³² <https://www.thetimes.co.uk/article/driverless-cars-taught-by-grand-theft-auto-mtzlv8rhk>

of consent and law. History shows that surveillance has negative societal effects, even if profitable for some.

2.2: When playing games under surveillance, beginners' luck is easily replicable

- 53 Beginner's' luck is defined as "the supposed phenomenon of novices experiencing disproportionate frequency of success or succeeding against an expert in a given activity. One would expect experts to outperform novices - when the opposite happens it is counter-intuitive, hence the need for a term to describe this phenomenon."⁹³³ There are other similar terms, such as "Hail Mary play"⁹³⁴ – a very remote chance where, while it probably won't work, there are no better options. So a player takes a chance and, every so often, it works. Humans celebrate, and go back to playing the game.
- 54 Without a way to build a human model of decision-making, AI developers simply feed inputs and outputs of games into a function which records every interaction, and aims towards maximised outputs. Given enough chaotic inputs, random actions, and computer power to run enough iterations, what works best in that game emerges from the chaos. This is beginner's luck in practice, functional only because of ubiquitous surveillance and linkage of cause and effect. An intelligence – whether human or artificial – with no idea what to actually do, makes completely blind random decisions. Sometimes, these turn out well. Hail Mary plays normally don't work, but they work often enough that if you have no other options, it is worth trying. The critical point being that, in simulators, and necessary for AIs, everything is recorded in minute detail.
- 55 While humans cannot rewind real world decisions that came out well – we cannot calculate exactly the mechanics of a throw, and then save it for future repetition, including predictions of intermediate moves to make such scenarios more likely to recur – that is precisely what a "deep learning" training does.⁹³⁵ This is the advantage of the machines, but it is entirely dependent upon the original inputs. While a primitive version of this is done by professional sports teams⁹³⁶ and TV sports commentators, even with their budgets, they are constrained by the level of input data. Human decision-making may be between alchemy and science, but applied AI *is* engineering. And engineering can be analysed.

⁹³³ https://en.wikipedia.org/wiki/Beginner%27s_luck

⁹³⁴ https://en.wikipedia.org/wiki/Hail_Mary_pass#In_other_fields

⁹³⁵ e.g. "engineering a tunnel so the ball hits the top of the screen": wired.co.uk/article/google-deepmind-atari

⁹³⁶ c.f. "Moneyball"

56 Run for the community of Go players to play against others, the KGS Go Server is a long-standing community service. It stores the games, and makes past games available in bulk to download. In public statements,⁹³⁷ DeepMind say that AlphaGo was trained on the games played there⁹³⁸.

57 In an interview with Wired magazine, “one of the creators of AlphaGo explained:⁹³⁹

“Although we have programmed this machine to play, we have no idea what moves it will come up with. **Its moves are an emergent phenomenon from the training. We just create the data sets and the training algorithms. But the moves it then comes up with are out of our hands—and much better than we, as Go players, could come up with.**” [emphasis added]

58 While the first sentence is likely entirely true, there is evidence that the part we have emphasised is in fact wrong⁹⁴⁰ – evidence cited by the people at DeepMind themselves. It appears that Google DeepMind, in the rush to self-promoting press releases, didn’t go back and check the training data... We did...

59 The type of move claimed as “an emergent phenomenon from the training”, and “much better than we, as Go players, could come up with” is, in fact, none of those things – it appears many times in the training dataset.⁹⁴¹ Even when we restricted our search to moves played at the

⁹³⁷ Behind a paywall: <http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html>, but summarised at https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol#AlphaGo

⁹³⁸ <https://www.gokgs.com>

⁹³⁹ <https://www.wired.com/2016/03/googles-ai-wins-pivotal-game-two-match-go-grandmaster/>

⁹⁴⁰ Some time has passed since the statement was made, so it is possible it is now disavowed - although we have seen no public evidence of that.

⁹⁴¹ Sample KGS .sgf files where this class of move, in the same area of the board, were played at this point of the game, not always by the same colour, since 2001, and before 2016 (odd-numbered games only) include: 2001-10-10-1.sgf, 2001-10-13-9.sgf, 2002-03-17-23.sgf, 2003-04-10-5.sgf, 2003-07-20-25.sgf, 2003-08-04-17.sgf, 2003-09-26-3.sgf, 2003-09-29-11.sgf, 2003-12-17-11.sgf, 2004-01-31-19.sgf, 2004-04-28-9.sgf, 2004-05-10-21.sgf, 2004-10-13-1.sgf, 2004-12-20-7.sgf, 2005-09-21-9.sgf, 2005-10-10-55.sgf, 2005-11-03-11.sgf, 2005-11-18-13.sgf, 2005-11-20-33.sgf, 2005-11-27-31.sgf, 2006-03-15-17.sgf, 2006-04-19-9.sgf, 2006-06-02-11.sgf, 2006-06-07-15.sgf, 2006-08-19-17.sgf, 2006-09-09-15.sgf, 2006-10-09-29.sgf, 2006-12-29-3.sgf, 2007-01-03-1.sgf, 2007-03-18-25.sgf, 2007-07-04-23.sgf, 2007-07-24-15.sgf, 2007-10-18-3.sgf, 2007-11-15-13.sgf, 2007-11-22-17.sgf, 2008-01-14-27.sgf, 2008-05-03-35.sgf, 2008-05-04-47.sgf, 2008-06-22-17.sgf, 2008-06-27-33.sgf, 2008-10-12-19.sgf, 2008-11-25-21.sgf, 2008-12-13-11.sgf, 2008-12-15-33.sgf, 2009-01-23-11.sgf, 2009-02-19-11.sgf, 2009-02-23-63.sgf, 2009-04-08-7.sgf, 2009-08-25-41.sgf, 2009-10-17-25.sgf, 2009-11-18-27.sgf, 2009-12-28-3.sgf, 2010-03-19-19.sgf, 2010-04-12-17.sgf, 2010-04-23-9.sgf, 2010-10-15-51.sgf, 2010-12-13-35.sgf, 2010-12-19-39.sgf, 2011-01-10-15.sgf, 2011-03-08-25.sgf, 2011-03-17-37.sgf, 2011-03-21-5.sgf, 2011-08-03-49.sgf, 2011-09-14-55.sgf, 2011-09-18-3.sgf, 2011-09-22-45.sgf, 2011-10-27-57.sgf, 2011-11-03-41.sgf, 2011-11-25-13.sgf, 2011-12-14-49.sgf, 2012-02-24-13.sgf, 2012-05-01-53.sgf, 2012-05-21-11.sgf, 2012-06-01-27.sgf, 2013-01-02-1.sgf, 2013-01-02-19.sgf,

same point in the game, that move had been played by many human Go players, who presumably didn't think they were doing anything special. They were just playing a move in a game,⁹⁴² possibly nearly 20 years ago.

- 60 Games have long contained "cheats" which allow a player who "discovers" them to achieve outcomes that others do not know are possible – this applies to AI's too. ⁹⁴³ Feeding all the KGS Go games to an AI meant it found a correlation that humans had missed, and, using its ability to measure precisely, it was able to find or encourage a scenario where it could 'see' opportunities humans would have not thought to look. AI is not magic.
- 61 Given what is now known about AlphaGo, and the progress that has been made in AI since it first emerged, it is now entirely possible for any individual or organisation – should they wish to commit enough computer time to the problem, to replicate its success. More generally, any AI trained solely on a non-proprietary dataset cannot be a proprietary AI.
- 62 As a result, public sector bodies, especially NHS bodies, should *not* sign exclusive AI contracts – but rather treat the new class of AI data processors as they do all other services and commodities contracted in from external experts. This will require procurement rules, and a statutory mandate for transparency to the citizen on how personal data is used.
- 63 With your bank statement, you have an evidence base for every change in your bank balance – you can go back and look at what happened. It gives you confidence in your bank that while you can hold them accountable, everyone else can do so too, and while you may not check it this month, enough other people probably do that it's ineffective to cheat. Whatever rules emerge, and given the nature of the UK press and public sector, if the citizen does not have an understanding of how their data is used, then there will be superstition and fear. If each citizen has a complete and honest accounting of how their data has been used, then, while they may not like the decisions taken,⁹⁴⁴ they should not fear anything is unknown.

2013-02-16-1.sgf, 2013-02-25-5.sgf, 2013-03-26-31.sgf, 2013-09-09-21.sgf, 2013-09-30-11.sgf, 2013-10-15-45.sgf, 2014-03-30-23.sgf, 2014-09-29-3.sgf, 2014-10-13-19.sgf, 2015-02-02-17.sgf, 2015-02-12-3.sgf, 2015-02-23-35.sgf, 2015-04-11-3.sgf, 2015-12-13-37.sgf. We have not checked every game for a more precise formulation - we lack the tools and capacity to do so.

⁹⁴² We thank a single unknown commentator on some Go message board for posting, at the peak of the AlphaGo hysteria, that they believed this to be true from their own experience. We found that post in 2016 when looking into the KGS archive. In writing this submission, we have failed to find that post again; we found the post via beginner's' luck, and fortunately, do not live in a society of ubiquitous surveillance.

⁹⁴³ <https://forums.frontier.co.uk/showthread.php?t=258662>

⁹⁴⁴ This is accounted for by the democratic processes of a country.

64 As a result of medConfidential’s work, in July 2017 such transparency and accountability was announced by HMG for most of the NHS, beginning in late 2018.⁹⁴⁵ Properly implemented, those steps will also serve confidence in AI well.

medConfidential

4 September 2017

⁹⁴⁵ <https://medconfidential.org/2017/medconfidential-response-to-the-governments-caldicott-3-response/>

medConfidential – Supplementary written evidence (AIC0244)

65 Paragraph 34 of our evidence said:

“At the time of writing, [26] Google DeepMind refuses to answer the question, “Did you feed the data to your AI?”

66 Since that evidence was published, DeepMind has published a blog post under the title “Why Doesn’t Streams Use AI?”, which says:⁹⁴⁶

“As well as the additional workload, it would have required us to effectively split our team into two to ensure that the Royal Free’s personally identified data (for Streams) and de-identified data (for research) were kept entirely separate. So we didn’t move forward with AI research, and nor did we sign the additional agreements with the Royal Free that would be required to do so. To this date, we have not done any research or AI development with the Royal Free.”

67 medConfidential takes such statements at face value.

68 We also take at face value a contradictory response from another entity which questioned why DeepMind would say that, given knowledge of what Google DeepMind did with the data at the time (i.e. GDM took actions consistent with their statements to the Health Research Authority in 2015⁹⁴⁷).

69 We entirely concur with the recently published 2017 Review from the National Data Guardian, Dame Fiona Caldicott, which says:⁹⁴⁸

“In summary, the goal should be a state of information governance in which the following proposition prevails: Organisations have no hiding places, the public have no surprises.”

70 Given the hiding places from which Google DeepMind choose to operate, it remains true today that correct governance did not occur in this case, and the public were unpleasantly surprised. There are few areas of work where it is more necessary or sensitive than an AI company processing medical records, whether they used AI or not.

⁹⁴⁶ <https://deepmind.com/blog/streams-and-ai/>

⁹⁴⁷ <https://techcrunch.com/2017/08/31/documents-detail-deepminds-plan-to-apply-ai-to-nhs-data-in-2015/>

⁹⁴⁸

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/666480/NDG_Progress_Report_FINAL.pdf

71 The NDG Review paragraph continues: “Statutory recognition would remove the hiding places and reduce the scope for public surprise”, and we are support that the National Data Guardian Bill has reached Committee stage in the House of Commons, and look forward to it receiving a warm and speedy reception in the House of Lords on its way to the statute book.

AI in the Data Protection Bill

72 Paragraph 4 of our evidence to the committee said:

“AI is not magic. AI bestows on its creators, users, and victims no capability that is not data processing. It may be novel data processing, it may be highly processing-intensive data processing, but it remains just data processing. We have laws for that.”

73 The Data Protection Bill will become that new law. The committee call for evidence closed before clauses 175-178 were added to the Bill. They allow the Secretary of State to create a “Framework for Data Processing by Government” (c175), which covers all data held by any public body, including the NHS (175(1)), and is both outside of the ICO’s jurisdiction (c178(5)) and under the control of Ministers (c175(4)), with courts bound by the framework (c176(7)), as are tribunals (c178(2)), and it only changes when required by international law (c177(4))⁹⁴⁹, and operates retroactively (c178(3)).

74 The Secretary of State mentioned is expected to be DCMS - which seems odd as DCMS is not known as a strong data processing Department, but it is explained by the Minister’s reply to Q191 in Committee which said (emphasis added):

Q191 Matt Hancock MP: **We think it will be resourced by civil servants reporting directly to Ministers. The office for AI is part of government. It is not independent. It is the team that will manage this policy development and architecture.**

...

I would say that we are the two lead departments on it, BEIS for the application and the wider economy through industrial strategy, and us for the AI sector itself and the digital strategy. **We have a joint unit because it naturally falls into both departments**, and, as you can see, we have an exceptional ministerial-level relationship.

Matt Hancock MP: **That insight is at the core of the need for the centre for data ethics and innovation.** The centre was proposed in the Conservative Party manifesto, because we, too, spotted that gap.

⁹⁴⁹ Brexit’s changes to jurisdiction of international law are ignored by the Bill.

Whenever any great new technology comes along, it is important that we harness the opportunities while mitigating the risks.

.... **We want to ensure that the adoption of AI is accompanied, and in some cases led, by a body similarly set up not just with technical experts who know what can be done but with ethicists who understand what should be done so that the gap between those two questions is not omitted. I am delighted that we have now been funded in the Budget in order to set it up.** It is incredibly important to ensure that society moves at the same pace as the technology, because this technology moves very fast.

- 75 The proposal the Minister suggested to the Committee is to create a quango, "reporting directly to ministers", "part of Government", explicitly stating "it is not independent. It is the team that will manage this policy development and architecture", "at the core of the need for the centre for data ethics and innovation".
- 76 It is therefore unclear why anyone would consider decisions made by this unit ethical - a Minister's job is to be political. It is all too easy to see elsewhere in the Data Protection Bill (e.g. Schedule 2 paragraph 4) how those two things are not only different, but may be incompatible. There are many things that are entirely lawful, but whether they are ethical is the subject of infinite debate.
- 77 The current Ministers championed their working relationship, but Ministers come and go. The law will remain for future Governments to use very differently with very different capabilities than available today.
- 78 There is a legitimate place for AI in Government. However, there has been no public debate on what that should look like. Instead, Government has chosen to legislate in haste, for a framework which will allow an AI to handle some data of the processing of the NCC1 form of DWP (the "rape form")⁹⁵⁰, or immigration choices (as a supplier suggested to the House of Commons Home Affairs Select Committee⁹⁵¹).
- 79 Perverse uses other than intended are a fundamental problem at the core of "AI safety". Perhaps clauses 175-178 should be removed and rethought, until Government offers substantive proposals and oversight for data processing by AI.
- 80 Committee has had next to no evidence on this topic (this evidence supplement now technically provides some) - the Committee's Call for

⁹⁵⁰ <https://www.gov.uk/government/publications/support-for-a-child-conceived-without-your-consent>

⁹⁵¹ <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/home-affairs-committee/home-office-delivery-of-brexit-immigration/written/73030.html>

Evidence had closed before Government laid the framework clauses. It would be unfortunate if Government chose to abuse your report and evidence for purposes other than you intend, by implying that no one objected when Ministers said quite clearly what they intended to do.

medConfidential

17 December 2017

The Medicines and Healthcare products Regulatory Agency (MHRA) – Written evidence (AIC0134)

The Medicines and Healthcare products Regulatory Agency (MHRA) is an executive agency of the Department of Health. The agency has 3 centres: the Clinical Practice Research Datalink (CPRD), a data research service that aims to improve public health by using anonymised NHS clinical data; the National Institute for Biological Standards and Control (NIBSC), a global leader in the standardisation and control of biological medicines; and the MHRA, the UK's regulator of medicines, medical devices and blood components for transfusion, responsible for ensuring their safety, quality and effectiveness.

Q6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

- There are a huge range of possible benefits to the UK health and social care sectors from the use of AI. It is already in use and will continue to grow. Current EU regulation states that all algorithms and apps that claim a medical purpose are medical devices and must comply with medical device directives. This is also true of AI. AI has great potential to transform the UK health and social care sectors but it is not without risk.
- The UK takes a risk-based approach to healthcare regulation. Regulation aims to ensure an acceptable level of risk, in a proportionate manner and without stifling innovation. The Agency provides guidance for industry to help in the classification of medical devices. It is also working closely with the Department of Health and other authorities to ensure that regulation develops to enable the UK health and social care sectors to meet this challenge.
- 'Big data' (including genomic, proteomic, imaging, epidemiological data) being generated in pharmaceutical companies is being utilised to aid the discovery and characterisation of potential medicines. Analysing these data sources individually and in combination is a fertile area for AI applications. As pharmaceutical companies increasingly utilise these technologies in biological medicine/vaccine development then, to ensure the continued safety and efficacy, these technologies are likely migrate into the associated governmental control testing.
- A further example of use is decision tree algorithms which offer diagnosis or treatment and work on large data sets with some element of continuous learning which may change patient treatment regimes on an on-going basis. These may have the ability to reduce face-to-face time with clinicians and can

be an important part of a care pathway and patient self-management pathway. However, for medical device algorithms of this type, the need to emphasise the requirement for transparency on how algorithms are continually tested, verified and clinically validated is paramount. It is important also that these processes are made understandable to clinicians to gain support and use.

- Further possible health uses of AI, many of which will qualify as medical devices are:
 - Diagnosis / prognosis of clinical conditions.
 - Use in developing evidence for medicines submissions and / or in clinical trials.
 - Product development of diagnostics, devices.
 - Use in combination systems e.g. a medicine with an app that could be constantly learning and changing action parameters or an AI natural language processing app used as an interface to control a physical device such as a robot.
 - Genomics / personalised medicine-selection of therapies; choice of medicine.
 - As part of vigilance / market / post-market surveillance of medicines and devices in identifying new signals in large databases of e.g. adverse incident data.
 - Real time prediction/detection/monitoring of pandemics/epidemics (Ebola, flu, etc).
 - Radiotherapy tumour segmentation.
 - Production control in the manufacturing process of medicines and devices.

Q7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- Data that have been generated within the public sector in relation to the delivery of public sector services should be used for public good. In the case of the UK health and social care sectors, patient data are collected based on confidentiality and for the explicit purpose of informing patient care and improving public health. It may be prudent to consider intellectual property rights, and patents for any AI products (for e.g. a predictive algorithm used to support clinical decision making) developed using public sector data and cost

implications of licensing these for subsequent applications within the public sector.

- A recent Wellcome Trust report explored UK public attitudes to commercial access to health data in a survey. The report did not focus on AI research but the findings are generalisable to AI. The survey found that in general, people were happy for their personal data to be used for research but many felt uncomfortable about sharing their data with commercial organisations, particularly if they were seen to be motivated by their own private interest with little or no public benefit. There was also a strong view that robust data governance systems needed to be in place to provide the public and patients, the assurance that their data would be used in a transparent and responsible way.
- It will be critical for health and social care services to engage with patients, their friends and relatives and the public more generally to explain to them why their data is needed, how it will be safeguarded, and how they will benefit from the new technologies and treatments it enables. A key lesson from past efforts to create patient databases, such as care.data, is that we need to do more to earn public trust.
- The report suggested four criteria that should be met before commercial organisations were provided access to their data: that the activity's outcome have a provable and sufficient public benefit, that the organisations undertaking the research could be trusted to have public interest at heart, that the data be anonymised (and risk of reidentification of an individual was minimal) and that there were robust safeguards in place including data governance and security measures.
- The report itself did not include specific recommendations on what such safeguards would involve or how they could be implemented. Possible safeguards could include nominating data custodians who control the release of anonymised / pseudonymised data based on the principles outlined in the Wellcome Trust report while ensuring transparency, scientific rigour and a favourable benefit / risk ratio.

CPRD case study

- The Clinical Practice Research Datalink (CPRD) is a dedicated UK Government initiative jointly supported by the MHRA and the National Institute for Health Research (NIHR) to provide electronic health record (EHR) data for public health studies. For more than 30 years, CPRD has been providing EHR data to enable high quality health research.
- CPRD data is used worldwide by regulators, academic researchers and industry to conduct public health research. Access to patient level data is provided for health research purposes only and is dependent on approval of a research protocol by the MHRA Independent Scientific Advisory Committee (ISAC) in accordance with data governance procedures and research ethics.

- CPRD’s data governance framework is not specific to AI based research but it is flexible and robust enough to encompass AI. This was reaffirmed in a recent stakeholder consultation workshop on machine learning organised by CPRD and including machine learning experts. The following excerpt from the workshop report illustrates this point:
- *"The overall view was that most machine learning methods did not pose fundamentally new risks or caveats as compared to classical statistical techniques; there may be slightly increased risks (due to the capability of machine learning to deal with an increased number of attributes) or they may present in slightly different guises, but essentially, concepts like sampling bias, internal validity and external validity applied to machine learning as well.*
- *As these risks and approaches to dealing with them are well recognised in epidemiology, existing guidelines on good practice in epidemiology could be adapted to propose standards for the conduct and reporting of machine learning.*
- *A key question put to stakeholders was whether the existing ISAC governance framework was fit-for-purpose in relation to machine learning research proposals or if it needed adapting for this purpose. The consensus was that there was no need for special guidance on assessing machine learning proposals and that the existing ISAC framework was robust enough to deal with these methods. It was acknowledged however that some further discussion was required to understand how a machine learning proposal should be written in a transparent way to enable reviewers to assess its merits using the ISAC review guidelines."*
- It is important to note however that some machine learning techniques like artificial neural networks pose an increased risk of hidden bias and work is underway to better understand and minimise the risks posed to public health by these biases both within CPRD and in the wider scientific community. Caveats relating to machine learning/artificial intelligence have also been highlighted in a paper published by the Information Commissioner’s Office (ICO):

"Machine learning itself may contain hidden bias. A common phrase used in the discussion of machine learning is "garbage in garbage out". Essentially, if the input data contains errors and inaccuracies, so will the output data. While supervised machine learning in particular often involves a pre-processing stage to improve the quality of the input data, the human-labelling of a training dataset can create a further opportunity for inaccuracies or bias to creep in. Hypothetically, a predictive model used in recruitment may achieve an overall accuracy rate of 90%, but this may be because it is 100% accurate for a majority population who make up 90% of applicants but wholly inaccurate for minority groups who make up the other 10%. It would be necessary to test for this and build in corrective measures."

(Clause 96, page 44)

- Therefore, providing appropriate safeguards (like validation and regulation) are in place, AI could enhance the development of healthcare applications including identification of novel drug targets.
- Key references:
 - CPRD Machine Learning Stakeholder Consultation Workshop Draft Report (2017).
 - Ipsos MORI (2016). The One-Way Mirror: Public attitudes to commercial access to health data. Wellcome Trust (available at: <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf>; accessed on: 16.08.2017)
 - ICO (2017). Big data, artificial intelligence, machine learning and data protection. (available at: <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>; accessed on: 05.09.2017)
- **Q9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?**
- Machine learning approaches primarily focus on developing prediction algorithms by 'learning' from data inputs. These approaches can be considered as mimicking human 'cognitive' functions i.e. 'learning' and therefore have been described by some as 'artificial intelligence'. Exact definitions and boundaries between the two terms have been widely debated but for this response, the two will be considered synonymous.
- Machine learning approaches incorporate an element of 'learning from errors'. There is a much lower tolerance for errors in healthcare that could place patients at risk of adverse outcomes. Therefore, it is important to emphasise transparency and validation for AI applications designed for healthcare.
- Not all algorithms employed in machine learning are opaque and these algorithms can be treated as many classical statistical techniques; the assumptions underpinning these methods and the associated limitations are well understood which enables their findings to be translated into practice in a sensible way.
- Some methods like neural networks which are associated with machine learning approaches, are not well understood (therefore the reference to 'black-box' algorithms). They can generate algorithms with extremely high predictive accuracy but do not allow the researcher to understand how that prediction was made.

- This means that it is more difficult to identify possible errors without additional validation tests. There is ongoing methodological work in the wider scientific community to address this limitation. For instance, even if the algorithm is not fully understood, it is possible to identify which features / attributes have factored predominantly in the prediction; this would allow researchers and reviewers to assess the validity of the algorithm.
- There may be issues around the auditability in case of failure. Safeguards around audit trails may be needed. AI outputs / predictions are meant to improve with increased input data and therefore AI systems should benefit from frequent updates. What if a diagnosis AI increasingly misdiagnoses cancer biopsies then how would that be investigated? If results from any given sample in the past cannot be recreated exactly because the AI system has evolved from its previous states then the ability to identify/correct potential causes is compromised. Furthermore, how does this affect accountability? It is important for AI, as any other algorithm that validation and verification of software integrity (and interoperability with software platforms) is maintained after any updates.
- The following views were expressed at the CPRD machine learning stakeholder workshop:
- *"Concerns were expressed about some machine learning methods like artificial neural networks (ANN) that are also sometimes referred to as deep learning. These methods have also been referred to as 'black box algorithms' as they employ latent (or hidden) variables in the analysis that may make it difficult for researchers and reviewers alike to assess whether there are any increased risks of inadvertent re-identification of patients or discrimination against certain population subgroups over and above classical statistical methods. These concerns could be mitigated by insisting on validation of models in other datasets or requiring researchers to demonstrate reproducibility of algorithms. It was noted that different analytical packages could yield different results so specification of software used was important; some benchmarking in terms of methods would also be prudent. Initiatives like 'OpenML' (<https://www.openml.org>) were highlighted as examples of how transparency could be promoted in machine learning by allowing for verification of algorithms by other researchers.*
- *Concerns were also expressed around the 'naivety' of some data scientists who believed that given large volumes of data, they could develop algorithms with better predictive performance than clinicians. Another concern was about researchers who wanted to use computing power to 'circumvent knowledge' or compensate for 'deficiencies in knowledge' (or theoretical understanding). It was reiterated that rationalisation of the choice of algorithms was crucial as it showed understanding of the methodological approach in general and the specific methods used.*

- *In relation to 'black box' algorithms like neural networks, it was felt that researchers should attempt to explain their models and initial model assumptions in a way that would be understandable to non-experts in machine learning. It was not acceptable for researchers to avoid justifying their choice of algorithms or explaining their models and model assumptions.*
- *Algorithms developed to support clinical decision making needed to be understandable by clinicians to engender trust in the algorithms. It was also important for end-users of algorithms to understand the limitations of the algorithms so that they could apply their clinical/personal judgement and overrule an algorithm prediction if it seemed counter-intuitive rather than blindly accepting algorithm outputs."*
- Key references:
 - CPRD Machine Learning Stakeholder Consultation Workshop Draft Report (2017).

Q 10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- As mentioned above, where AI meets the definition of a medical device it must be regulated including being CE marked under the EU Directives, which are due to be replaced by new EU Medical Device / IVD Regulations in Member States in May 2020 and 2022 respectively. This is to ensure acceptable levels of patient safety and public health. There is much work underway across relevant authorities in government and with industry to determine how the elements of regulation can be applied to these systems to ensure they are acceptably safe and workable.
- If AI systems are to be regulated then the training/input data utilised is integral to the system as a whole. This is especially true in the heterogeneous, 'big data' medical research field.
- Machine learning incorporates an element of 'learning from errors'. There is a much lower tolerance for errors in healthcare that could place patients at risk of adverse outcomes. Therefore, it is also important to carefully consider transparency and ongoing validation for AI applications designed for healthcare.
- Classical statistical techniques and the associated limitations are well understood. However, some methods associated with machine learning approaches are not well understood (hence the term 'black-box' algorithms). These can generate algorithms with extremely high predictive accuracy but do not allow understanding of how that prediction was made. This means that it is more difficult to identify possible errors without additional validation tests.

- There is much discussion on how verification and validation of something that constantly changes (a key feature of continuous learning) and its clinical impact is achieved especially for 'deep-learning / black box' algorithms. How often should AI / machine learning devices be re-tested once in use? A certain amount of risk needs to be accepted as one develops and learns. What level is acceptable? The manufacturer will need to provide evidence that processes are in place that ensure validity and risk are not compromised or increased when changes are made to the algorithm, including:
 - Determine the level and methods of pre-testing of AI algorithms prior to use e.g. On test/synthetic data or 'sandbox' testing to ensure a certain safety level.
 - Provide verification of test results (possibly by independent sources).
 - Having evidence to ensure frequency of on- going testing during the life cycle of the product maintains safety levels and clinical validity.
- Government has an important but challenging role in regulating the use of AI in medical devices to ensure a balance between innovation and patient safety. The biggest challenge will be in adapting regulation to address the individual features of fast changing AI algorithms. This is important because, while there are many potential healthcare benefits from AI, these technologies are not without considerable potential risks. This is especially true within health and social care settings.

6 September 2016

Microsoft – Written evidence (AIC0149)

David A. Heiner Strategic Policy Advisor*

INTRODUCTION

1. Microsoft welcomes the opportunity to respond to the Lords Select Committee enquiry into Artificial Intelligence (AI). Microsoft believes that the power of “intelligence” through computers can bring great benefits to society—enabling important advances in education, healthcare, transportation, sustainability, economic efficiency and many other areas. Realising the full economic potential of AI will require an appropriate policy and regulatory framework, and we applaud the Committee’s timely efforts to solicit public dialogue on these subjects.
2. AI may greatly boost economic growth. Recent research by Accenture estimated that AI could add £140 billion to the UK economy by 2034, increasing the annual economic growth rate from 2.5% to 3.9% by 2035, and boost labour productivity by 25% across all sectors, including Britain’s strong pharmaceutical and aerospace industries.
3. Microsoft is investing heavily in the research and development of AI technologies because we believe that AI is central to the digital transformation that is at the heart of economic development. In August 2017, for example, our researchers announced that they had developed an AI system that can recognise words in a conversation even more accurately than most people—an industry milestone. AI technologies such as this can enable people to communicate with each other while speaking different languages, in applications such as Skype Translator which enables real-time voice-to-voice translation, or Presentation Translator, which enables captions to be displayed automatically in any of 60 languages during a presentation. Such an application is also of benefit for those who are hearing impaired.
4. It is no exaggeration to say that AI has the potential to save lives. For example, the biological computation group at our Microsoft Research Lab in Cambridge is working at the intersection of machine learning, computer-aided design, and biology to pioneer new approaches to challenges such as treating cancer. Researchers are also collaborating with biologists, radiologists, and other medical experts to use advanced computational methods to understand the behaviour of cells and their interaction, which will help to “debug” an individual’s cancer and provide personalised treatment.
5. AI can also enable us to better address environmental concerns. Microsoft recently launched AI for Earth in London—a new initiative dedicated to sustainability challenges, including agriculture, water, biodiversity, and climate change. We plan to invest up to £1.5m in qualified initiatives and offer NGOs and other groups working on environmental issues access to AI tools, services, and technical support.

6. Every significant technological advance has raised a range of societal issues, and AI is no exception. Governments, civil society, industry and researchers must be thoughtful as AI is developed and deployed so as to bring about the greatest benefit for all. This includes addressing the possibility of job displacements, and developing best practices to ensure that AI systems are safe to use, respect privacy, are transparent and fair. In this submission, we provide suggestions on how to address these concerns.

SHAPING AI DEVELOPMENT AND PUBLIC PERCEPTION

7. Microsoft’s vision for AI is straightforward: we aim to amplify human ingenuity with intelligent technology. We believe we can augment human intelligence through advances in computer vision, speech recognition, natural language processing, and machine learning generally. In this regard, we believe the term “Artificial Intelligence” does not adequately describe the technology and innovation as there is little that is “artificial” about it. *People* developed the powerful microprocessors, data storage capacity and machine learning techniques that increasingly enable computers to perform tasks that in the past only humans could perform. A better term might be “computational intelligence”—intelligence that can help address some of society’s greatest challenges.
8. One of the ways we aim to augment human intelligence is to make AI available to all through a range of technological programs. Initiatives such as Microsoft Cognitive Services, the Microsoft Cognitive Toolkit, the Bot Framework, and Azure Machine Learning enable software developers, enterprises and others to draw upon advanced AI techniques developed by Microsoft in building their own computing solutions.
9. The human-centred approach to AI that Microsoft envisions can only be realised if relevant stakeholders from industry, government, civil society and the research community collaborate on the development of shared principles to shape the use of AI technologies.
10. Microsoft’s CEO, Satya Nadella, shared some initial thoughts on what these may be in order to start this dialogue. We believe that AI should:
 1. Be designed to assist humanity;
 2. Be transparent;
 3. Maximise efficiencies without destroying the dignity of people;
 4. Be designed for privacy;
 5. Have algorithmic accountability so that humans can undo unintended harm;
 6. Guard against bias.

Complementing the above are key considerations for everyone developing, deploying and using these technologies:

1. Empathy;
2. Education (knowledge *and* skills);
3. Creativity;
4. Judgment and accountability.

11. A first step towards articulating a common vision was taken in September 2016, when Microsoft, together with Amazon, DeepMind/Google, Facebook and IBM, launched the Partnership on AI (PAI) with a mission to “study and formulate best practices on AI technologies, to advance the public’s understanding of AI, and to serve as an open platform for discussion and engagement about AI and its influences on people and society.” Since then, various companies, civil society organisations, researchers, and others from Europe and Asia have joined PAI to help formulate a set of AI principles.

WORKFORCE TRAINING

12. Since the dawn of the Industrial Revolution, technological advances have eliminated or greatly reduced the number of jobs in some categories, while creating entirely new categories that would have been hard to imagine just a few decades earlier. In the UK, for example, more than 20% of the workforce was employed in the agricultural sector at the turn of the 20th century, while today, thanks to mechanisation, less than one percent works in agricultural roles. However, that small group produces vastly more food than their predecessors a century ago and overall unemployment in the UK is near a 40 year low at 4.4%.
13. Although AI has greatly advanced in the past few years, most AI systems are limited to what researchers refer to as “narrow” AI. These are systems that can perform particular tasks well (for example, play a game, describe an image, or respond to a simple inquiry), but these systems do not have any kind of artificial “general” intelligence—the kind that could replace a person who performs a range of tasks and whose job requires the exercise of significant judgment.
14. Predicting the future is notoriously difficult, and there appears to be no consensus on whether AI will create more jobs than it replaces. A recent report in the US pointed out that there is neither sufficient data nor metrics to provide longitudinal insights into workforce trends and predicted demands. How AI, and technology generally, impacts the organisations where it is deployed and the workforce and economy-at-large is not yet well understood; nor how AI might play a role in mitigating these challenges. Measurements and tools can help policymakers develop better public policies, as well as enable ongoing monitoring of workforce, skills demands, and training strategies. Microsoft is working with other organisations, including funding academic researchers, to address some of these questions.
15. Although the likely overall effect of AI on jobs is uncertain, there is general agreement that advances in AI will lead to the reorganisation of businesses, requiring the workforce to acquire new skills and remain adaptable to re-skill or up-skill their abilities throughout their lives. McKinsey recently estimated that nearly one third of the tasks in a little over half the jobs that people have today can be fully automated. Given the inevitability of efficient automation, it will be essential that everyone develops “soft” skills—those that are unlikely

to be replaced by AI, including social skills, creativity, planning and organising.

16. We should prepare students for the future world of work by providing a solid foundation in science, technology, engineering, and mathematics (STEM) education, and helping them to develop the soft skills noted above. Students, especially young women, should be encouraged to develop technology skills, such as computer and data science, and adaptive skills, such as communication and creative thinking. Microsoft has long been a strong proponent of the need for more STEM education in schools. In the UK, we have provided significant funding since 2014, investing in teacher professional development, guidance for school leaders, research into the pedagogy of computer science at school, and supported the development of the computer science GCSE. Working with the Department for Education Microsoft provided funding for the QuickStart Computing training programme for teachers, the largest such teacher training programme in over 20 years. It is vital that teachers continue to be supported in a way that enables them to deliver the new curriculum in the most effective way possible.
17. Despite these and other efforts, in a year when China and India each produced 300,000 computer science graduates, the UK produced just 7,000. Microsoft welcomes the development of new 'T-Levels' to strengthen technical education in the UK and would like to see its implementation expedited in order to minimise the technical skills gap deficit in the UK. The Council for Science and Technology has also outlined three key principles that the Government should pursue to capitalise on the industrial potential of robotics, automation and artificial intelligence (RAAI). These include increasing the number of RAAI-related facilities in the UK, identifying ways to for greater collaboration in the industry, and developing advanced skills and research capability, all of which Microsoft would welcome.
18. Alongside investment in education, the skills needed to thrive in AI-driven economies are rapidly evolving and will require people to continually update their skills throughout their careers. A World Economic Forum [report](#) estimated that about half of the subject knowledge acquired during the first year of a four-year technical degree is outdated by the time students graduate. This makes life-long efforts at skills development essential, especially valuable middle-skills which can be obtained through certification, vocational training, continuing education programmes, and apprenticeships. In January 2017 Microsoft UK announced a commitment to train for free 30,000 civil servants in digital skills and launched a Cloud Skills Initiative, which will train 500,000 people in the UK in advanced cloud technology skills by 2020.
19. New tools, such as those provided by LinkedIn (a Microsoft company), can help to predict and identify needed skills, enable effective training for inclusive growth, and better connect available skills and opportunities. Over time, this data can be used to construct analyses such as the [LinkedIn Economic Graph](#) to provide transparency into the supply of and demand for

skills, especially if they can be combined with other data held by governments on local demographics and businesses. In the U.S., Microsoft is working with local governments, companies, non-profits, and other organisations to better understand the impact of technology disruption, and the potential for using new tools and data to develop solutions. An example of this is our philanthropic work with the Markle Foundation in the U.S. to expand data-driven approaches to connecting workers and businesses.

20. We urge policy stakeholders to address concerns about jobs by focusing on the development of the skills of the workforce, both technological skills and soft skills. The U.K. Government-commissioned Taylor Review into modern employment practices highlighted that increasing automation would have an impact on the U.K. labour market, but stressed the ability of automation to enhance the working experience, rather than rendering it redundant. Microsoft agrees that the labour market must remain dynamic to reap the benefits of advances in RAAI and to provide jobs complement RAAI.
21. Government, business, educators, and other interested stakeholders should foster innovative solutions to these challenges, work to better understand the impact of AI on jobs, and help create an appropriate policy framework for the economic transformation enabled by AI to be as inclusive and benefit as many people as possible. We look forward to reviewing the Government's response to the Taylor Review and how it addresses the impact of new technologies on the labour market.

DESIGNING AI TO EARN TRUST

22. As AI plays an increasing role in mediating people's lives online and offline, appropriate design, economic and social choices will be essential to ensuring that the technologies will be deemed *trustworthy* by individuals and society at large—and be regarded as *respectful* and *inclusive*. The computational power and learning capabilities of machines must be coupled with the sensitivity and emotional intelligence of humans. Fulfilling the potential of AI requires IQ *and* EQ.
23. AI systems should be developed in accord with universal, timeless values. We believe it is especially important that AI systems be designed to be safe for all users, fair to all, transparent, privacy protective and inclusive. We touch on each of these below.
24. **Safety:** AI-based systems must demonstrate that they can be depended upon to operate correctly, reliably and safely, consistently over time, both under normal operating environment as well as when under attack from bad actors. A key requirement is that an AI system be trained with large amounts of data and that the data be representative of the fact patterns on which the system will operate. Second, AI systems must be trained to understand our intended meaning, rather than to take what we say literally. (When we instruct a self-driving car to "take me to Heathrow as quickly as possible," we

probably do not mean “race down the M4 at 100 miles per hour.”) Third, and very importantly, AI systems must be designed to expect the unexpected: When expected conditions no longer apply, AI systems should fail gracefully, such as by ceding control to a person, while providing appropriate information so that the user can take control effectively. More broadly, we need more research into human-robot interaction, including how AI systems can signal and communicate with people in settings of shared responsibility.

25. **Fairness:** It is imperative that AI systems be designed to treat all people fairly. This is important because absent sufficient care an AI system could in fact treat people unfairly, even while being wrapped in the aura of scientific precision. Modelling human behaviour with numbers can lead to valuable insights, but is challenging. For example, a researcher developing a system to make hiring recommendations may want to use a job applicant’s reliability as an input, but this probably cannot be measured directly. The researcher might consider using the applicant’s credit score as a proxy, but that could reflect life circumstances other than the applicant’s reliability in paying back loans. AI researchers must be especially sensitive to the data they use to train their systems. AI systems learn based on the data they are fed, and if that data reflects the biases of the individuals or organisations that collect or curate the data, the cultural biases and behaviour of society at large, or is incomplete, biases may be learned, reinforced, and, in some cases, even amplified by the resulting AI models.
26. There are three key steps we can take to help address these challenges. First, we should redouble efforts to attract a diverse workforce to the computer industry, and AI in particular. Second, we should encourage and fund research into the development of data analytics techniques to identify when an AI system may be returning unfair results and to show how to fix that. The Fairness, Accountability and Transparency in Machine Learning community is already making strides in these areas. Third, we should develop guidelines for AI researchers to aid them in developing systems that treat all people fairly. We are working on developing such guidelines internally at Microsoft, and together with the industry through PAI.
27. **Transparency:** As AI-based systems are increasingly used to make decisions that affect people’s lives in important ways, people naturally want to understand how these systems operate, and why they make the recommendations that they do. Enabling transparency of AI systems can be challenging due to their complexity and the fact that recommendations are largely a function of an understanding of massive amounts of data, which computers excel at, but people do not. (And access to data about people often cannot be provided, given privacy considerations.) Researchers are developing a number of promising techniques to help provide transparency, such as developing simpler systems that closely mimic the recommendations of more accurate systems yet are easier to understand, and systems that enable people to vary various inputs to see the effect on system recommendations. Microsoft is working with PAI to develop best practices to

enable useful transparency. Such best practices will likely include an explanation of the system objectives, the data sets used to train the system during development and in deployment, selection criteria for the algorithms, the system components and their interactions, testing and validation of the system, and risk mitigation considerations.

28. **Privacy:** AI systems that concern people will need access to data about people to function. However, people will not make their data available to AI systems if they don't believe that their data will be used carefully and securely, and according to their interest or the interest of the community at large. Strong data protection laws such as the General Data Protection Regulation (GDPR) are important to enable AI to flourish. Data protection laws should be applied with sensitivity not only to privacy but also to the benefits to all that AI can enable, if sufficient data is made available. Context will be important. For example, an individual's name in a company's internal employee directory would not typically be considered sensitive and generally requires less privacy protection than the same name appearing on a "black list" related to credit ratings. To take another example, an individual's sexual orientation would typically be considered sensitive, but this should not restrict the processing of pension records, which include the name and gender of the partner of an employee, thereby revealing the sexual orientation of that individual. Processing of sensitive information generally should be enabled where such processing is in the public interest, including in enabling accessible applications, in the field of employment law, for monitoring and alert purposes, or for the prevention or control of communicable diseases and other serious threats to health. In fact, processing of such data will often be essential to enable AI researchers to test whether their systems are inadvertently discriminating on the basis of such categories.
29. To help address privacy concerns while enabling data use, we should encourage research into, and the deployment of, data de-identification and anonymisation techniques. AI researchers need vast amounts of data about people in order to design systems that concern people, but they often do not need to know the identity of particular people. Since these techniques may not be able to guarantee anonymity, it may be helpful to complement use of these techniques with legal prohibitions on efforts to "re-identify" people in the data sets. This is akin to putting a lock on the door to our homes, and passing laws to prohibit breaking into our homes. While security is not guaranteed, the lock, backed by the force of law, is useful.
30. We must also ensure that data continues to flow freely between the UK, EU, and other countries post-Brexit. We welcome the Government's commitment to implement the GDPR through the Data Protection Bill. However, once the UK leaves the EU it will no longer automatically be a part of the EU-US Privacy Shield. Consideration must also be given to how data flows can continue uninterrupted with both the US and the EU. This, and the issue of adequacy between UK and EU data protection regulations, will need continued

focus to ensure that the UK remains a world leader in new technologies like AI.

31. **Inclusion:** AI technologies should be developed in a way that benefits and empowers everyone. AI can be empowering for the more than 1 billion people around the world with disabilities—increasing access to education, employment, government services, and social opportunities. AI can help people become or remain productive and independent, regardless of abilities and age. Applications, such as those embedded in Office 365 and [Seeing AI](#), a free Microsoft app on the iPhone for the visually impaired, can help individuals with vision impairment to engage more fully in professional and social contexts.
32. AI can help governments ensure that everyone has equal access to information, services, the political process and jobs. Increased workforce participation by people with disabilities can lead to increased incomes and higher GDPs. To achieve these benefits, governments should focus their policy-making on: procurement—making accessibility a criterion for public sector procurement of ICT; standards—leveraging international harmonised standards; e-government—adopting policies that mandate accessibility for government information and e-government services; and inclusive education—integrating accessible technology into classrooms and learning solutions. The UK has implemented the EU accessibility standard EN301 549 on accessibility and procurement and ratified the United Nations Convention on the Rights of Persons with Disabilities, which requires that countries adopt legislation and take steps to promote the rights of people with disabilities in the use of information and communications technology, education, and employment. AI is an asset that can be leveraged to enable innovative offerings to deliver such services.

POLICY RECOMMENDATIONS

33. As AI is still at a nascent stage of development, a continuing collaboration between government, business, civil society and academic researchers is essential to shaping the technology and realising its benefits. Working together, we can identify and prioritise issues of societal importance as AI continues to evolve, enable sharing of best practices and motivate further research and development of solutions as new issues emerge. Policy discussion should prioritise broad development and deployment of AI across different sectors and continued AI innovation, encouraging outcomes that are aligned with the vision of human-centred AI.
34. We offer a few suggestions for UK policy makers to consider in creating an enabling policy framework for AI:
 - Continue to convene dialogues between government, business, researchers, civil society and other interested stakeholders on how AI can be shaped to maximise its potential and mitigate its risks, including adoption of practical guiding principles to encourage development of human-centred AI;

- Encourage sharing and promulgating of best practices in development and deployment of human-centred AI, through industry-led organisations such as PAI;
- Stimulate the development and deployment of AI across all sectors and business of all sizes by:
 - Incentivising small and medium enterprises to leverage AI, as they are key in addressing income stagnation amongst less affluent households,
 - Promoting use of AI in the public sector, enabling more informed policy decisions and more personalised services,
 - Encouraging innovative applications of AI to address public and societal challenges,
 - Promoting use of AI to empower underserved communities and those with disabilities;
- Implement the GDPR through the Data Protection Bill and agree to a successor to the Privacy Shield post Brexit to support development and uptake of AI; encourage use of data anonymisation techniques;
- Invest in skills training initiatives for people at all stages of the job continuum;
- Fund short- and long-term multi-disciplinary research and development of human-centred AI, including those that address the timeless values raised above, and how AI can be used to provide additional insights into socio-economic issues that may be caused by deployment of the technology. The research should consider areas that private industry is unlikely to pursue (e.g., public health, urban development, smart communities, social welfare, criminal justice, environmental sustainability, national security), and longer term transformational impacts of AI on society.
- Develop shared public data sets and environments for AI training and testing, to enable broader experimentation with AI and comparisons of alternative solutions to address ethical concerns.

35. AI has the potential to transform and improve every aspect of our lives. We look forward to contributing to the UK government's ongoing efforts to develop an enabling policy framework to realise this vision.

6 September 2017

Dr. Zdenek Moravcik – Written evidence (AIC0019)

Dr. Zdenek Moravcik, inventor of „human brain algorithms“(i.e. artificial general intelligence)

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Answer: Given the term "artificial intelligence" equals to "software simulation of functions of human brain" then such a software simulation of human brain is currently available. There is no need to develop/invent/research this topic any further. I am the author and inventor of this brain simulation which is sometimes also called artificial general intelligence. I am willing to transfer my invention to any country (including UK) which is interested and willing to make use of this technology. Please contact me.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes.

You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

*Answer: My brain software simulation inserted into a humanoid machine makes this humanoid robot into a "**universal robot**" capable of doing tasks so far only humans could do. This makes most of the human activities **redundant** (like work in the factories) but does NOT make human force obsolete in any way. There will be enough new jobs for the humans to do. There is no need for irrational fears. And there is no need to slow things down!*

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Answer: Everyone is gaining as everyone will be working less and everyone will become more in the age of "universal robots". The whole society will be granted capabilities to finance activities that were up to now not possible (like mass education of humanity).

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Answer: Governments worldwide are failing to both control and address the issue. Instead of inquiring about persons/groups who claim to have seriously something to do with the issue, irrational news coverage is being spread in the media resembling primitive propaganda/conspiracy. Governments need to directly talk to persons like me who know the best about what such a technology can/will do.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

Answer: Universal robots can generally take over any job in the factory so all industry sectors will be affected.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence?

How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

Answer: Not sure what this question objectively means but as long as governments are handling this new technology in an irresponsible way it is highly expected that wide public has no chance of understanding this new technology and its impacts correctly. The results of such irrational situation will not be good.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so – called 'black boxing') acceptable? When should it not be permissible?

*Answer: Question irrelevant as principles of how human brain is working (i.e. processing information) are **fully** understood by me!*

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Answer: Governments worldwide are failing to cope with this new coming technology. Governments should at first place INQUIRE about people like me who claim that simulation of human brain is already available. Only me as an inventor of human brain simulation can give government reliable information about what is going to come in the next years, what this technology is all about... Ignoring it is a serious political error.

*Government must also **support project of building first intelligent robots** based on my simulation of human brain as this is nothing a single company (no*

matter how big it is) or single researcher can do reliably. No support from government and/or slowing things down is again serious political error!

21 August 2017

Dr Ian Morgan and Brian Joyce – Written evidence (AIC0179)

The views expressed in this report are our own personal views and in no way represent O2 or Telefónica.

Summary

- The terms artificial intelligence and machine learning have been conflated over time, and although this is mainly a matter of semantics the terms do evoke different images; it is suggested that (at least, in the short to medium term) it is machine learning approaches that are most relevant to society.
- It is posited that there will not be a revolution in the job market, more of an evolution in the form of ever increased automation of white collar jobs, primarily in locations where employment is expensive.
- Topics of education, especially STEM, and specifically statistics and programming will become more important to remain competitive.
- Large, post-digital companies are the primary winners in machine learning due to ready access to their own datasets and the ability to compete in the job market where skillsets are limited.
- The majority of sectors will be impacted; essentially anywhere data can be collected on physical (e.g. gas turbines) or digital assets (e.g. transactions or previous law cases) and customer behaviour.
- In terms of regulation, it is argued that the input and output of a process should be regulated (where necessary) and that this is already mostly in place. There may be new regulation required around newer markets for the purposes of defining liability for the insurance industry.

Definition of artificial intelligence (AI) and machine learning (ML)

1. The definition of AI has been somewhat conflated with machine learning, where the terms are often used synonymously. From my perspective, AI can be seen as a separate area of research where the future aim of such research is to have a general intelligence that cannot be easily distinguished from human intelligence. From a purist view, I would argue that a generalised AI and the solution to 'NP-complete' problems is still far-off, although advancements in reinforcement learning and computational power have perhaps brought the possibility of that particular future slightly closer.
2. Self-driving cars, machine to text translation, and game playing are impressive advances in different domains, supported by advances in probabilistic reasoning, neural networks and computing scale, however are all specific to each vertical and could not be considered to be generally intelligent. Pragmatically, however, the term AI could also be applied wherever a system is seen to be doing something 'smart' or 'intelligently'. It is also likely that the definition of AI changes over time, where a task like

identifying a person in an image may have been considered intelligent in earlier years.

3. Machine learning therefore, is the set of algorithms and processes that can be applied for AI, as well as for more 'mundane', business related tasks, such as credit risk calculation or automatic system monitoring. These algorithms and processes typically learn through repeated applications to the same task, or large sets of historical data.
4. I would suggest that it is these algorithms, and consequently machine learning, that is the focus of the research group, as it is the analysis of large amounts of data and automation of jobs that is more relevant to today's society.

Questions

A.1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1. ML, AI and data science over the past 5 years is potentially reaching peak exposure in the media and public consciousness, where significant investment has been spent and delivery will shortly be expected; and in some cases the promise of ML will fall down under scrutiny. Despite this, ML and AI are here to stay, and where applied correctly will mean that companies can do more with less cost and potentially fewer employees.
2. Cost reductions to date in various industries have come from areas including consolidation of services from outsourcing. That is likely to change to some extent, where menial tasks are looking to be automated, to increase oversight and reduce mistakes in a variety of fields. This is both customer facing; to increase response times and reduce load on customer support staff in the form of chatbots and (for example) Amazon Alexa skills, as well as automating internal processes; for example, automatic classification of credit applications to identify fraud, automated network monitoring to identify, diagnose and even predict anomalous scenarios (e.g. a network failure).
3. In the next 5 years, I do not believe that there will be a revolution in the jobs market, more an evolution as was seen with the introduction of computers, and more recently the internet, where the available jobs changed especially in the technology sector. I would argue that ML is there to support human action, and typically, human intervention will still be required with some processes; just perhaps less of it.

4. In the longer term, where companies have become more accepting of automatic computer intervention this may remove the requirement of having humans to action the decisions; for example the self-driving car market. Here, as in many industries, it is regulation that can protect, hinder or enable that intervention. As we have seen with public transport (e.g. the DLR compared to the tube) it is not necessarily technological but societal expectations and requirements that are the limiting factor in wholesale automation.

B.2. Is the current level of excitement which surrounds artificial intelligence warranted?

Answered in other questions

C.2. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

1. It is our belief that machine learning will initially be retrofitted into existing markets to streamline process; whether that is reducing energy consumption on the network, reducing customer waiting times when contacting customer service, or enabling smaller teams looking at credit applications. As a result, consumers will increasingly have to interact with automation, both in their professional and home lives. It also has the potential to disenfranchise those with no understanding of the underlying systems, especially when no care has been made to relate the reasoning behind decision making processes or targeting rationale. There is a risk that more and more employment will move toward software and automation, and more so in the western hemisphere where employees are a more expensive resource.
2. In terms of preparing individuals for the introduction of wider scale machine learning initiatives, STEM provision in the UK has been inadequate for some time, both at the school and undergraduate level, especially in the area of machine learning and probability theory. Even the tuition of classical statistics is typically left until university, and the application of such techniques left to irrelevant examples with no industry focus.
3. There is significantly more information freely available than in previous years; blogs, online teaching courses as well as open source which has been readily adopted by large and small companies alike, which will assist the learning

curve to some extent, however there should be more focus on computer science, statistics and machine learning while still at school.

4. From a retraining perspective, further education has been extensively cut, and course fees at universities are typically excessive for mature students, reducing applications by around 50% over the last 5 years, so financial support for those who wanted to retrain would be invaluable.
5. Although not strictly related to machine learning and artificial intelligence, a by-product is that companies are now collecting far larger amounts of data, even where the data is not directly related to their primary business, and this does need to be carefully controlled; both from a consumer perspective (who should be reassured that their data is safe, and is not shared), as well as from hindering business from developing a competitive edge on a global stage.

D.4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

1. Most companies who have or rely on large quantities of data will at some stage gain from the use of machine learning. This gain can take many forms. For example, if you have a large customer base, ML and targeted algorithms will make it easier to understand the nuances of certain customer groups making advertising campaigns more personalised and hopefully more effective.
2. In her book, 'Weapons of Math Destruction' Cathy O'Neil points out several groups of people primarily in the USA (however there are UK examples too) who have suffered as a result of this use of big data. These normally include people who are economically challenged as well as having little or no education.
3. One of the possible ways to mitigate disparities or injustices that may come about as a result of ML and big data is to make the algorithms transparent or the people/organisations developing them accountable. If, for example, someone were to lose their job as a result of an algorithm applied by their employer, that person should have the right to understand the detail behind the algorithm in order to mount a credible defence.
4. AI is limited by the quality of the data, the scope of implementation and appropriateness of a particular algorithm or modelling technique. It is rare for all of these to be perfect. Data is usually messy and inconsistent, scope of implementation usually gets expanded past its original purpose and ML algorithms by their very nature should always be learning.

E.5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

Answered in question 4.

F.6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

1. Above all, any sector that collects data in some format can apply machine learning in their business. Clearly, businesses that have some data policy already in place can maximise their use of data as they will readily have access to their own information. Somewhat contrary to intuition, it is the large, pre-digital companies that may have the hardest time utilising machine learning in a well-defined way, due to the size of the company and the many disparate means of holding their own data. Furthermore, dependent upon the type of data there are also restrictions that have to be worked with; e.g. analysing personal demographic information as opposed to using anonymised information or generalities. It is certainly larger companies, with larger datasets, and multiple areas of possibility that can gain the most benefit or advantage from machine learning. As a relatively small application, for example, better fraud detection can disproportionately affect thousands or millions of transactions.
2. Consequently the aims for implementing machine learning may be different for different companies; smaller businesses might make the machine learning core their unique selling proposition, whereas larger companies might implement machine learning in a variety of different areas to support their primary business function.
3. Somewhat inevitably, larger companies in any sector will have an easier time due to the ability to compete in the job market and retain suitable individuals, due to both being able to offer higher salaries, as well as having a wider variety of problems to work on with a higher impact value.

G.7. How can the data-based monopolies of some large corporations, and the ‘winnertakes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Somewhat answered in question 6.

H.8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

1. There are, of course, ethical implications not just with machine learning but with the collection of large amounts of publically and privately available data, both explicit (e.g. comments on a public forum, reviews on Amazon, a Facebook profile) and implicit (websites visited, minutes spent on a particular product page, browsing device type, or cellular location information). It turns out that the digital economy is significantly more quantifiable than the offline version, as everything is recorded, and consequently those companies in the digital economy who do not use their own data are at a disadvantage to those that do as inevitably it is possible to be more targeted with fewer resources.
2. This is a similar story in the electoral process, where parties must do as much as possible with capped resources (at least in the UK). In theory this is a good thing, as it is possible to define what matters to people and why in increasing detail. The ability to misuse the information is not a new thing, and despite ever more sophisticated classification and profiling techniques we can see from recent examples that there is still much more work to be done to categorise people and their behaviour.
3. There are some safeguards in place in terms of data collection, storage and usage, and to some extent it is both public awareness of these activities as well as the assumption that corporations have some consideration for their own reputation that must to some extent be relied upon. Possibly the largest trawlers of information are the security services, which may implicitly give the green light to others that wide scale profiling and collection of data is reasonable.
4. In terms of new markets and automation, for example autonomous cars, we have given some examples in question 10.

I.9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

Answered in question 10.

J.10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

1. There are some in the AI and machine learning domain, most famously Elon Musk, who have suggested that AI itself should be regulated.
2. Despite having laudable aims, it is notoriously difficult to regulate or even patent software. For example, the United States (known for having a particularly relaxed patent system), has not allowed patents for software instructions or software itself, and after 7 years of litigation, Google has been found not to infringe Oracle copyright on the Java programming language. Even if regulation was appropriate, it could lead to fire-fighting against firms who adapt and change a particular feature to sidestep any rules.
3. It is certainly possible to regulate the process of software development, around documentation, source control and testing, which is present in both telecommunications and aviation among other industries in the form of ISO 9000, however this is on a company wide basis rather than on individual software projects; it would also be unreasonable to expect small companies to sign up to such heavy regulation.
4. As is common with existing regulation, it should be both the input and output to a process that is regulated. In new markets, this may require new regulation; for example the scenario of whether a self-driving car should protect either the occupants or other members of the public in the case of an accident as well as liability for insurance purposes. In existing markets much of the regulation around data storage and outputs from a process are already in place; in telecommunications (among other industry) the impending GDPR regulations.
5. Consequently there might be scope for machine learning specific process regulation (as the existing ISO 9000 might not be useful in such cases), a code that could be used by providers of life critical systems (aviation, military and so forth) as this might enable further use of machine learning in this domain. However, it is my thought that application of such a code would be of limited use in the commercial arena.

K.11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

We were not aware of policy in this area, apart from two reports released by the Obama administration;
<https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy>.

The summary pragmatically focuses on the impacts of automation, and summarises that it is likely to be important to train and educate the population, further invest in AI, and aid workers to transition through the changes in the economy to reduce any disproportionate economic impact.

Background

Ian Morgan completed a BSc in artificial intelligence and psychology at the University of Birmingham later graduating from the University of Portsmouth with a PhD in computational techniques applied to condition monitoring, and has since worked in a variety of companies in a technological capacity, including General Electric and O2 Telefonica. He has had a number of conference and journal submissions on machine learning (ML) techniques applied in industry and now works in the O2 Telefonica Labs, on data focussed projects.

Brian Joyce has a background in dealing with big data for top tier law firms. He has worked as a programmer for several years in O2 and currently holds the position of Information Engineer where he works on several big data projects involving machine learning.

6 September 2017

Dr Sarah Morley and Dr David Lawrence – Written evidence (AIC0036)

Call for Evidence- Artificial Intelligence

Dr. Sarah Morley and Dr. David R. Lawrence

Importance of Regulating AI

1. There is an existing and urgent need for the Government to develop new policies and regulations that address the emergence of new types of artificial Intelligence (AI). AI will have different levels or degrees of consciousness; the Government will need to create legal definitions for this consciousness in order to distinguish between the different legal responsibilities that will inevitably arise for the AI itself and for those who develop and operate AI.
2. The regulation of AI is paramount when considering the increasing use of these technologies in our daily lives and their increasing consciousness. A survey of the 100 most cited academics writing on AI suggests an expectation that machines will be developed "that can carry out most human professions at least as well as a typical human,"⁹⁵² with 90 percent confidence, by 2070, and with 50 percent confidence by 2050. While it must be stressed that this is merely educated speculation, the prototypes and experimental robots extant today are more than impressive. The componentry and systems exist (though for now they are yet to be united in one machine) to emulate proprioception, tactility,⁹⁵³ visual processing and object recognition, walking and running⁹⁵⁴—even on rough terrain and at high speeds⁹⁵⁵—and many more elements of human biology, even the high-speed recognition, analysis, and reaction needed to play table tennis.⁹⁵⁶ Robots have long been a feature of the workforce; for example, in the automotive manufacturing industry, but are now in a position to start taking more subtle, customer-facing jobs. ASIMO, Honda's famous walking robot, has acted as a receptionist,⁹⁵⁷ and has acted

⁹⁵² Müller VC, Bostrom N. Future progress in artificial intelligence: A survey of expert opinion. In: Müller VC, ed. *Fundamental Issues of Artificial Intelligence*. Cham, Springer; 2016:553–71.

⁹⁵³ Syntouch- Biotac. *syntouchllc.com*. 2016. Available at:

<http://www.syntouchllc.com/Products/BioTac/>. Accessed July 14, 2016.

⁹⁵⁴ ASIMO – The Honda Worldwide ASIMO Site. *World.honda.com*. 2016. Available at:

<http://world.honda.com/ASIMO/>. Accessed July 14, 2016.

⁹⁵⁵ Raibert M, Blankespoor K, Nelson G, Playter R. BigDog, the Rough-Terrain Quaduped Robot.

Boston Dynamics. 2008. Available at: http://www.bostondynamics.com/img/BigDog_IFAC_Apr-8-2008.pdf. Accessed July 14, 2016.

⁹⁵⁶ Raibert M, Blankespoor K, Nelson G, Playter R. BigDog, the Rough-Terrain Quaduped Robot.

Boston Dynamics. 2008. Available at: http://www.bostondynamics.com/img/BigDog_IFAC_Apr-8-2008.pdf. Accessed July 14, 2016.

⁹⁵⁷ Humanoid robot gets job as receptionist. *New Scientist*. 2005. Available at:

<https://www.newscientist.com/article/dn8456-humanoid-robot-gets-job-as-receptionist/>. Accessed July 14, 2016.

intelligently in concert with other ASIMOs as a team of office assistants.⁹⁵⁸ Many industries live in fear of the encroachment of automation,⁹⁵⁹ and robots are even expected to move into the “educated professions” such as law and medicine.⁹⁶⁰ Robotics and artificially intelligent systems are not a future issue, rather, they are very much an integral and essential aspect of modern society; and they will continue to become ever more so as development continues across the world.

3. When we consider the potential stakes; smart systems that could upend our society, or the birth of AGI that could think and reason like a human, with wants and needs and perhaps moral rights of its own, there is probably good reason to want to ‘get in front’ of these challenges; but it does not necessarily follow that we *should* do so, or do so unilaterally. To try to control or limit the development of robotics and AI may prevent responsible and conscientious parties from doing so, but it will not stop others. With the potential impacts so significant, it seems that the sensible approach would be to ensure that freedom to act rests in the hands of those most able (or those likely to be so) to do so appropriately and with consideration for consequences. Guidelines and regulations that attempt to control technologies after the fact are rarely great successes, and with one as ephemeral as an AI (of any type) it will be all the more difficult. Furthermore, with regard to AI the balancing act of scientific freedom and the preservation of the status quo is a futile endeavour- AI will, no doubt, be the greatest technological challenge to our society, and has already fundamentally altered how we live.

Role of the Government

4. There is an urgent need for the Government to produce policies and regulations that address the emergence of AI and the involvement of corporations in their creation and operation. Moreover, as AI will have different levels of consciousness the Government will need to consider how this should affect its regulation. For example the Government will need to form legal definitions for this consciousness in order to distinguish between the different legal responsibilities that will inevitably arise for the AI itself and for those who develop and operate AI.

⁹⁵⁸ The World's Most Advanced Humanoid Robot. *Asimo by Honda*. 2016. Available at: http://asimo.honda.com/news/honda-develops-intelligence-technologies-enabling-multiple-asimo-robots-to-work-together-in-coordination/newsarticle_0073/. Accessed February 16, 2017.

⁹⁵⁹ Why robots are coming for US service jobs. *Financial Times*. 2016. Available at: <http://www.ft.com/cms/s/0/cb4c93c4-0566-11e6-a70d-4e39ac32c284.html#axzz4DNsK7QYF>. Accessed July 14, 2016.

⁹⁶⁰ Meltzer T. Robot doctors, online lawyers and automated architects: the future of the professions?. *The Guardian*. 2014. Available at: <https://www.theguardian.com/technology/2014/jun/15/robot-doctors-online-lawyers-automated-architects-future-professions-jobs-technology>. Accessed July 14, 2016.

5. The Government should therefore play a particular role in determining:
 - I. Legal definitions to determine the different consciousness and moral status of AI
 - II. The legal status of AI: Should AI be granted legal personhood?
 - III. The responsibility to AI: Who is responsible for the creation, lifespan and ultimate fate of AI. If the answer is the company who produced the AI, to what extent should they be liable?

Points two and three are likely to have different implications depending on the consciousness-derived moral status of the AI in question (hence they should be subsidiary to point one). These points are expanded upon in the following sections.

i. Consciousness and moral status of AI

6. As robotics have advanced, so too has the development of AI, in concert with the abovementioned and as a field in its own right. There are a number of subfields, each immensely complex, working toward elements of human-level intelligence. For example, a true, conscious AI would need to be able to perceive and understand information;⁹⁶¹ to learn;⁹⁶² to process language;⁹⁶³ to plan ahead and anticipate (and thus visualize itself in time);⁹⁶⁴ to possess “knowledge representation”⁹⁶⁵ or the ability to retain, parse, and apply the astronomically high number of discrete facts that we take for granted, and be able to use this information to reason; to possess subjectivity; and many, many more elements. A number of projects exist attempting to develop and integrate one or more of these elements into “artificial brains,” using modeled or biological neural networks and other technologies; including Cyc,⁹⁶⁶ an ongoing 32 year attempt to collect and incorporate a vast database of

⁹⁶¹ Russell S, Norvig P. *Artificial Intelligence A Modern Approach*. 2nd ed. New Jersey: Prentice Hall; 2003. at 537–81, 863–98.

⁹⁶² Langley P. The changing science of machine learning. *Machine Learning* 2011;82(3):275–9.

⁹⁶³ Cambria E, White B. Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine* 2014;9(2):48–57.

⁹⁶⁴ *Op cit.* 10 at 375–459.

⁹⁶⁵ *Op cit.* 10 at 320–63.

⁹⁶⁶ Knowledge modeling and machine reasoning environment capable of addressing the most challenging problems in industry, government, and academia. *Cycorp: Home of Smarter Solutions*. 2016. Available at: <http://www.cyc.com/>. Accessed July 14, 2016; The word: Common sense. *New Scientist*. 2006. Available at: <https://www.newscientist.com/article/mg19025471.700-the-word-common-sense/>. Accessed July 14, 2016. I thank John Harris for informing me of this fascinating endeavor.

“common-sense” knowledge in a practical ontology, to enable reasoning. There is also the Google Brain,⁹⁶⁷ a “deep learning” project focused on giving the AI access to Google’s vast troves of data and allowing it to begin to parse things for itself; for example, the Brain, when given access to *Youtube.com*, learned unprompted to recognize human faces, and showed a partiality to videos of cats.⁹⁶⁸ A third project, the well-known Blue Brain, has successfully modelled 37,000,000 synapses of a rat’s sensory cortex⁹⁶⁹ in an attempt to understand the “circuitry.”

7. It is imperative that the Government defines when an AI is both conscious and unconscious. This is because the different statuses of AI should have implications for the regulations that follow such as legal responsibility. For example, if the AI is deemed to be conscious regulations should reflect on whether the appropriate mechanisms for shutting down or “killing” the technology should be different from that of unconscious AI.
8. The present authors are currently undertaking research to consider these future technological developments and suggest practical legal definitions for the status of both conscious and unconscious AI, in service of later developing and providing proposals for appropriate regulation for the responsible development, operation, and disposal of the technologies. By way illustrating un-consciousness, we might consider that an intelligence of a type which surpasses our own raw cognitive processing power might warrant being called ‘super’, as it could, in a narrow sense, outperform us. But this type of AI is not likely to be conscious. This type of AI is the one which presently exists- albeit probably without yet qualifying as ‘super’. We can see examples in many AI which we utilise as individuals every day- from simple algorithms used by streaming television services such as Netflix which recommend shows based on your viewing history;⁹⁷⁰ to stock market trading programs;⁹⁷¹ to the complex Bayesian systems which operate autopiloting in aircraft and autonomous cars.⁹⁷² These are all ‘expert systems’⁹⁷³ or ‘applied’ AI

⁹⁶⁷ Hernandez D. The Man Behind the Google Brain: Andrew Ng and the Quest for the New AI. *WIRED*. 2013. Available at: <http://www.wired.com/2013/05/neuro-artificial-intelligence/>. Accessed July 14, 2016.

⁹⁶⁸ Google’s Artificial Brain Learns to Find Cat Videos. *WIRED*. 2012. Available at: <http://www.wired.com/2012/06/google-x-neural-network>. Accessed July 14, 2016.

⁹⁶⁹ Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and simulation of neocortical microcircuitry. *Cell* 2015;163(2):456–92.

⁹⁷⁰ Gomez-Uribe CA, Hunt N. The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*. 2016 6;4:13.

⁹⁷¹ Dymova L, Sevastjanov P, Kaczmarek K. A Forex trading expert system based on a new approach to the rule-base evidential reasoning. *Expert Systems with Applications*. 2016 Jun 1;51:1-3.

⁹⁷² Zhu, W; Miao, J; Hu, J; and Qing, L. Vehicle detection in driving simulation using extreme learning machine. *Neurocomputing* 2014 128: 160–165.

⁹⁷³ Cuddy C. Expert systems: The technology of knowledge management and decision making for the 21st century. *Library Journal*. 2002 127;16:82.

(sometimes known as 'weak' AI⁹⁷⁴)- based on the combination of a knowledge base and an inference engine. In effect, the system is pre-programmed to recognise data and to respond in a certain manner- so for instance an autonomous car might detect a sudden obstacle ahead and another vehicle pulling alongside, infer the risk of collision, and would be able to choose to swerve the opposite way. These systems are not making decisions in the manner of a human, using reasoning and intuition to consider cause and effect, but are instead applying their own type of first-order logical rules,⁹⁷⁵ which might at a very simple level be summed up as 'if X, then Y'.

9. If AI is shown to be legally conscious (having considered the legal definitions) does the AI have capacity to take on responsibilities? Because consciousness does not equal competence. The answer to this question will impact whether the AI should be given a legal personality and who is ultimately responsible for the AI. Current tests of capacity and competence from medical law can be used to test the AI on these matters. For example, *Gillick v West Norfolk & Wisbeck Area Health Authority [1986]* might be used to determine an AI's competence.

ii. Should AI be granted Legal Personality?

10. Recent proposals by committees of the European Parliament, the White House, and the House of Commons⁹⁷⁶ have suggested, among other things, the institution of corporate personality for extant 'expert systems' and autonomous robots. These proposals may not constitute an appropriate regime as they fail to address the subsequent technological development of full conscious beings, or the comparable implications of synthetic genomic design.
11. In undertaking the enormous task of regulating AI, the Government should firstly consider whether AI should be eligible to be accorded legal personality. The decision to award legal status to AI will have many ramifications for legal responsibility and for issues such as legal liability. Additionally, the conscious status of AI will need to be considered when deciding on this point. If AI are not awarded legal personality then the Government will need to decide who takes legal responsibility for these technologies, be it the developers

⁹⁷⁴ Searle, J.R. (1980) 'Minds, brains, and programs', *Behavioral and Brain Sciences* 1980 3: 3, pp. 417-57

⁹⁷⁵ Forgy, C. Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem *Artificial Intelligence*. 19;1: 17-37.

⁹⁷⁶ European Parliament Committee on Legal Affairs. (2016) *Draft Report With Recommendations To The Commission On Civil Law Rules On Robotics (2015/2103(INL))*. Brussels.

House of Commons Science and Technology Committee. (2016) *Report on Robotics and Artificial Intelligence*. London, HC145

National Science and Technology Council Committee on Technology. (2016) *Preparing For The Future Of Artificial Intelligence*. Washington D.C.: Executive Office of the President

(companies) or the owners. For example if a self-driving car crashes and causes injury to a third party, who will be responsible for paying the damages - the developers or the owner? It may be that the developer will be liable if there has been a fault with the AI machinery/programming but otherwise the owner should insure themselves against liability like any other car. In this instance the Government can amend current regulations to ensure owners of AI are insured against any losses they may suffer because of the AI. Criminal liability however may be more difficult to establish if the AI is not granted a legal personality.

iii. Legal Responsibility and Company Law

12. It seems likely that AI will be the product of public corporations and in particular multinational corporations. The main source of regulation for these corporations derives from company law. Company law here is to be understood to incorporate not only company law in the traditional sense (Companies Act 2006) but also other regulatory mechanisms that control the behaviour of companies such as criminal sanctions, civil remedies, governance codes etc.
13. Currently, there are no company regulations which specifically address the development and operation of AI. This includes the ethical and safe advancement and destruction of AI. For instance, as the law stands Directors are not required to consider whether AI should have a right to life, to liberty, or to self-ownership; nor to the impacts its existence and operations may have. There is no requirement for any such project to be disposed of in a responsible manner, taking into consideration that closure may involve the “killing” of the AI, or what the effects of an incomplete cessation of activity may be. Furthermore, if AI is determined to be conscious but not competent should companies be legally responsible for the AI until they can be proven to possess legal capacity?
14. How heavily corporations should be involved in deciding on these, often sensitive, matters will need to be considered by the Government. We would advise that companies should be regulated to some extent on these matters in order to protect society and the AI itself. We have already seen so-called ‘racist’ and ‘sexist’ AI resulting from bias implicit in coding by human agents, unintentional though it may have been.⁹⁷⁷
15. If companies are left unregulated in this area there is a further risk that AI will be affected by the specific drivers of companies (profit), and in particular of public companies (shareholder primacy and short-term profit maximisation). Are the traditional drivers of companies appropriate for the

⁹⁷⁷ Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. *Science*. 2017 Apr 14;356(6334):183-6.

development of any morally significant technological development? We would answer no.

16. This poses the question as to whether company law can, or should, be the primary means of regulating AI, and by extension their potential wide-ranging societal impacts. We would answer that there is certainly potential for AI to be regulated by current company law regulations. For example the Companies Act 2006 could impose specific duties on directors to develop, operate and dispose of AI in an ethical manner. The UK Corporate Governance Code could also be utilised to include specific guidance on these matters.

Conclusion

17. AI systems are pervasive, and are involved in almost everything that utilises digital automation. They are, in effect, so immersed in the fabric of our society that they *are* that society. It may well be that humanity could continue without applied AI, as we managed for many millennia, but it is certain that we could not operate in the same way as we do today. Nor could we enjoy the many benefits of these systems that we take for granted. Scientific progression in these fields, and its 'trickle down' into the smallest parts of our lives, has fundamentally altered the human experience. This has been a great benefit to those fortunate enough to enjoy it- and it is a great argument in favour of having the freedom to do so. However the influence of these systems, this irreversible interweaving of science and society, leaves us at a crossroads. Further integration of weak AI into our lives, or the pursuit of 'strong'⁹⁷⁸ or 'general'⁹⁷⁹ AI (that can go beyond problem solving into human-level cognition) through the free practice of science, is likely to cause more direct changes to who and what we are. Our place in the hierarchy of beings, even our relative position as the pinnacle of moral status could be forever altered.

18. As the stewards of scientific progress, we are beholden to all parties- both to existing persons, and to the beings we may create through AI research. The risks and fears surrounding AI are purely our problems to solve, or to prevent from arising through careful design and the implementation of appropriate regulation and policy to govern their development. This work is presently beginning- already bodies within nations likely to drive the research and technologies in question are exploring the challenges and proposing their own means of addressing them. Reports such as the White House National Science and Technology Council Committee on Technology's *Preparing For The Future*

⁹⁷⁸ Kurzweil, R. *The Singularity is Near* New York: Viking Press 2005

⁹⁷⁹ Newell A, Simon HA. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*. 1976. 19;3:113-26.

Of Artificial Intelligence, the UK House of Commons' Science and Technology Committee *Report on Robotics and Artificial Intelligence*, and the European Parliament's *Draft Report With Recommendations To The Commission On Civil Law Rules On Robotics* all emerged at the end of 2016, though it should be said that none of these documents are definite regulatory roadmaps. They do however aim to provide a basis for controlling the integration of AI and our lives- to bridge the gap between science and society in a controlled manner. Whether the suggestions will be effective is yet to be seen, but the fact these documents exist is a promising start. What we must ensure, though, is that we consider reality- whether advanced technological development is permitted or tightly controlled, there will always be the chance that it is developed in secret and beyond regulatory reach. We would therefore suggest that the Government does play a role in regulating fundamental issues to ensure that AI is developed and operated both safely and ethically, whilst still allowing innovation in science.

19. We propose that this role primarily consists in the first instance of approaching the three key points outlined in this document, i.e. to agree legal definitions and standards by which to measure the moral status of an AI, to thus determine whether a given AI is eligible for legal personhood, and to determine and enforce responsibility of creators towards any new AI person and in the production of new AI. These will provide a logical and well-founded basis for future legislation able to cope with the advent of developed, conscious intelligences.

30 August 2017

National Data Guardian for Health and Care – Written evidence (AIC0143)

About the National Data Guardian for Health and Care

1. The National Data Guardian for Health and Care (NDG) advises and challenges the health and care system to help ensure that citizens' confidential information is safeguarded securely and used properly.
2. Dame Fiona Caldicott was appointed as the first NDG by the Secretary of State for Health, Jeremy Hunt, in November 2014.
3. Dame Fiona believes it is important to build trust in the use of data across health and social care and in her role as the NDG is guided by three main principles:
 - encouraging clinicians and other members of care teams to share information to enable joined-up care, better diagnosis and treatment
 - ensuring there are no surprises to the citizen about how their health and care data is being used and that they are given a choice about this
 - building a dialogue with the public about how we all wish information to be used
4. Although sponsored by the Department of Health, the NDG operates independently, representing the interests of patients and the public. The NDG also appoints an independent group of experts – the NDG panel – to advise and support this work.
5. More information is available on the NDG webpages on GOV.UK: <https://www.gov.uk/government/organisations/national-data-guardian/about>

NDG interest in the inquiry

6. The NDG role is to advise and challenge the health and care system to help ensure that citizens' confidential information is safeguarded securely and used properly.
7. Dame Fiona's interest in this inquiry is primarily around the way that patient data⁹⁸⁰ might be used to develop and advance artificial intelligence and the extent to which this is done in a way that safeguards

⁹⁸⁰ Patient data is used in this submission to cover data collected from publically funded health and adult social care services

confidentiality, engages the public in a dialogue about how data is used and provides individuals with appropriate choice about this.

8. Dame Fiona and her advisory panel are planning to undertake some work to consider and better understand the implications of artificial intelligence for patient data. If the Lords Select Committee on Artificial Intelligence would find it useful at the oral evidence stage to engage with the further thinking that the NDG has been undertaking on this issue, she would be pleased to be of service to the committee.
9. This response focuses on the questions that the NDG is currently best able to address, namely 3, 5 and 8.

Responses

Question 3: How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

10. Artificial intelligence offers significant potential to the health and care system to improve the quality of care and outcomes for patients and service users. In order to achieve these benefits, it will sometimes be necessary to use patient data to develop and advance artificial intelligence tools.
11. In September 2015, the NDG was asked by the Secretary of State for Health to undertake a review of data security, consent and opt-outs.
12. The review was published in July 2016⁹⁸¹. It found that there are very low levels of public awareness and understanding of how patient data used. Patients and the public are generally not aware of how patient data is used and shared within and between NHS organisations, let alone how external organisations such as universities and technology companies might use patient data.
13. As NDG, Dame Fiona has said there should “no surprises” for the individual about who has had access to health and care information about them. This rests on an idea of reasonable expectations: citizens feel surprised where

⁹⁸¹ <https://www.gov.uk/government/publications/review-of-data-security-consent-and-opt-outs>

personally identifiable data about them is used in ways that depart significantly from uses that they accept or expect. Reasonable expectations are not static, but may shift over time as individuals' understanding and acceptance of data flows changes.

14. New technologies may create surprises where the pace of change in the purposes or scale of data flows significantly outpaces the pace of change of citizens' expectations. Public understanding of the way that patient data is used and might be used has not kept pace with the rapid acceleration of technology, including artificial intelligence.

15. The NDG believes that unaddressed, this understanding gap may lead to a diminution in public trust. In the absence of information, understanding and trust, anxiety may grow about whether patient data, which many individuals regard to be deeply personal, is being treated with the respect they would want and expect.

16. One of the key recommendations of the NDG Review was *"The case for data sharing still needs to be made to the public, and all health, social care, research and public organisations should share responsibility for making that case."*

17. The NDG therefore strongly advocates public engagement and transparency around the way that patient data is used. The need for this applies equally to the use of patient data in artificial intelligence, if not more so given the novelty of the technology to many members of the public.

Question 5: Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

18. With regards to the use of patient data to develop and advance artificial intelligence, the NDG firmly believes that efforts should be made to improve public understanding and engagement, for reasons explained in the response to question 3.

19. There may be lessons to be learned from work to engage and inform patients and the public about the way patient data is used in genetic and genomic medicine and science, another area where a key challenge has

been communicating to individual patients and to the wider public how data is used in a new and rapidly developing area.

Question 8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

20. Artificial intelligence raises challenge for the “no surprises” approach as the pace of change is likely to be rapid, and is also likely to accelerate over time. In order to ensure that innovation can still take place in a way which appropriately respects the reasonable expectations and wishes of the public, the NDG believes it is important that early consideration is given to what kind of governance arrangements will be suitable. Questions about social value, governance and ethics should be asked early and often. It might also be helpful to draw on approaches from outside health and social care, where attempts, both failed and successful, have been made over the past 25 years to build and maintain public trust in the face of disruptive innovations such as GM Crops, nanotechnology, and stem cell usage.
21. In relation to privacy, it is important to acknowledge that while sometimes it will be sufficient to use anonymised or synthetic data to develop and advance artificial intelligence tools, sometimes personally identifiable data will be needed.
22. Where personally identifiable patient data is required for the development of artificial intelligence tools, this also raises the question of what the legal basis would be. It will be important that where personally identifiable patient data is being used to develop and advance artificial intelligence tools, that there is clarity about the way the data is being used and where consent or another legal basis is appropriate.
23. With regards to anonymised information, the NDG Review heard that the public is broadly content for their anonymised information to be used where there is a clear health and social care purpose. Other research, such

as that carried out by the Wellcome Trust⁹⁸² and that undertaken by Connected Health Cities⁹⁸³, indicates that public support for the use of patient data, including anonymised patient data, is contingent upon there being a perceived public benefit as opposed to simply commercial gain.

6 September 2017

⁹⁸² <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-summary-wellcome-mar16.pdf>

⁹⁸³ https://www.connectedhealthcities.org/wp-content/uploads/2016/08/CHC-juries-report-Feb-2017_2.pdf

Professor John Naughton – Written evidence (AIC0144)

I am a Senior Research Fellow in the Centre for Research in the Arts, Social Sciences and Humanities (CRASH) at the University of Cambridge where I am co-director of two research projects, one on 'Conspiracy and Democracy', the other on 'Technology and Democracy'. I am also Emeritus Professor of the Public Understanding of Technology at the Open University and the technology columnist of the *Observer* newspaper. In the 1990s I published a history of the Internet.⁹⁸⁴ My most recent book -- from *Gutenberg to Zuckerberg: what you really need to know about the Internet* — is published by Quercus.

One of the problems with digital technology is that public discourse of it is often deluged with hype. If we take the long view, we see this as a recurring pattern: it happens in waves. First there is excitement prompted by some unexpected developments. These lead to wild extrapolations of the possibilities apparently opened up by the breakthrough, followed by commercial and industrial interest, investment by companies and venture-capitalists and — occasionally — real and effective deployments of new or improved products and services enabled by the technological breakthrough. This recurring pattern is usefully visualised in the Gartner Hype Cycle⁹⁸⁵ which places particular developments on a sentiment curve which starts with a 'trigger' (the initial breakthrough), followed by a frenzied increase in interest which culminates in a 'peak of inflated expectations'. This is followed by a precipitous decline which bottoms out in a 'trough of disillusionment', after which there is a slow crawl up a 'slope of enlightenment' and an eventual 'plateau of productivity'. Some (perhaps most) innovations never make it to the plateau. And for some that do the time elapsed between trigger and deployment may be years or even decades.

AI has been through several of these cycles⁹⁸⁶, and until recently few of the supposed breakthroughs ever made it much beyond the peak of inflated expectations. The current excitement about the field seems justified, in the sense that it is more likely to endure. This is due to a number of factors: (i) advances in machine learning; the availability of vastly increased processing power; Big Data; improved algorithms; and technical breakthroughs in neural networks; and (ii) the involvement of a number of large digital corporations with deep pockets which are not only giant attractors for highly-qualified and talented engineers and computer scientists but also have products and services which can benefit greatly from incorporating AI in them.

⁹⁸⁴ *A Brief History of the Future: the Origins of the Internet*, Weidenfeld, 1999.

⁹⁸⁵ https://en.wikipedia.org/wiki/Hype_cycle

⁹⁸⁶ https://en.wikipedia.org/wiki/AI_winter

One of the reasons for the so-called 'AI winters' of the past⁹⁸⁷ was that most of the research in the field was publicly funded, and relatively little was funded by corporations. Now the position is reversed.

It's important to distinguish between (i) 'strong AI' (more properly called Artificial General Intelligence and often dubbed 'superintelligence') — i.e. artificial intelligence where the machine's intellectual capability is functionally equal to a human's⁹⁸⁸ — and (ii) 'weak AI', i.e. "non-sentient artificial intelligence that is focused on one narrow task".⁹⁸⁹ Weak AI is what we have now, and it is largely a combination of machine-learning, Big Data and powerful algorithms.

Much of the public discourse about AI is media-driven and focused on supposed fears about the existential threats to humanity that would be posed by 'superintelligent' machines. This debate may be of interest to philosophers and tabloid editors, but at the present time it is a distraction from the important policy questions posed by the weak-AI that we already have. Various surveys have shown that very few of the established experts in the field believe that superintelligence is anything other than a very distant prospect. In one survey, 25% of respondents did not believe that it would *ever* be achieved.⁹⁹⁰

Accordingly, the focus of the Select Committee should be on the applications of weak-AI that are already embedded in devices and online services, and on the developments of that technology that are already in the pipeline and visible in prototype form.

An old joke in the AI community is that "AI is stuff that we cannot do yet". But the moment it gets implemented as a product or a service then it is no longer regarded as AI. The five global digital companies — Amazon, Alphabet (Google's holding company), Apple, Facebook and Microsoft — are already at that stage. Google's CEO, for example, describes the company's strategy in terms of "AI first" or "AI everywhere".⁹⁹¹ Similar rhetoric is now heard from the leading executives of the other digital giants. In practice this means that some combination of machine learning, Big Data analytics and algorithmic decision-making is already deeply embedded in the goods and services that they offer.

Machine learning — which essentially means computers having the ability to learn things without being explicitly programmed — is a core technology in this field. In many applications — for example spam detection, copyright

⁹⁸⁷ https://en.wikipedia.org/wiki/Lighthill_report

⁹⁸⁸ <https://www.ocf.berkeley.edu/~arihuang/academic/research/strongai3.html>

⁹⁸⁹ https://en.wikipedia.org/wiki/Weak_AI

⁹⁹⁰ <https://www.technologyreview.com/s/602410/no-the-experts-dont-think-superintelligent-ai-is-a-threat-to-humanity/>

⁹⁹¹ <https://www.fastcompany.com/3065420/at-sundar-pichais-google-ai-is-everything-and-everywhe>

protection and recommendation systems — it is effective and useful. In other kinds of applications — for example ‘predictive policing’ or decision-making on prison paroles or automated price-setting — it is controversial, and many many of the current concerns about algorithmic decision-making focus on: their opacity, the difficulty of accessing the logic behind their judgements; and questions of accountability (i.e. who is responsible if an algorithm makes a decision that has adverse effects on people?) The fear is that societies are moving towards what one leading scholar calls ‘the Black Box society’⁹⁹² — a world in which machines increasingly implement rules based on logics which are proprietary and incomprehensible to most people, especially those whose lives may be affected most by them.

Powerful new technologies invariably spark utopian and dystopian hopes and fears, and AI is no exception. If — as I believe — fears about ‘superintelligence’ are misplaced, concerns about the implications of ubiquitous weak-AI are not. On the contrary: recent events in the UK (and also in the US) suggest that a significant portion of the electorate feels powerless and excluded. Technologies that reinforce those perceptions are likely to increase this polarisation. It is therefore important that an appropriate regulatory environment is developed so that the advantages of AI technology can be realised without increasing polarisation and distrust.

It will be argued that if we place a brake on innovation in AI then other nations will overtake and outpace us — with consequences for both economic well-being and national security. The level of investment in AI research in China, for example, suggests that such fears may not be entirely groundless. A major study in the US has concluded that existing capabilities of AI technology (i.e. weak-AI) have “significant potential for national security”. Machine-learning could enable high degrees of automation in labour-intensive activities such as satellite imagery analysis and cyber defence.⁹⁹³ The report goes on to argue that “Future progress in AI has the potential to be a transformative national security technology, on a par with nuclear weapons, aircraft, computers, and biotech.”

The challenge, therefore, is to design regulatory regimes that provide reasonable safeguards for society while not unduly constraining the pace of disruptive innovation. This won't be easy. A possible approach would be to agree a set of general principles which would inform the formulation — and evolution — of a regulatory framework.

⁹⁹² See Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2016.

⁹⁹³ Greg Allen and Daniel Chan, *Artificial Intelligence and National Security: a study on behalf of Dr. Jason Matheny, Director of the U.S. Intelligence Advanced Research Projects Activity (IARPA)*, Belfer Center, Harvard Kennedy School, July 2017.

One set of such principles proposed by a leading AI researcher — Oren Etzioni⁹⁹⁴ —takes its inspiration from Asimov’s famous Three Laws of Robotics. Etzioni’s principles are: (1) an A.I. system must be subject to the full gamut of laws that apply to its human operator; (2) an A.I. system must clearly disclose that it is not human; and (3) an A.I. system cannot retain or disclose confidential information without explicit approval from the source of that information.

One major area of concern about the advent of weak-AI is its possible impact on employment, especially white-collar (‘middle class’) employment. A celebrated study by Frey and Osborne⁹⁹⁵ created a stir with an analysis that suggested that up to 47% of the 700-odd job categories used by the US Bureau of Labor might be vulnerable to automation by current AI technology. Much of the subsequent discussion of this and other studies focused on the fact that the jobs now supposedly at risk are white- rather than blue-collar ones. One study⁹⁹⁶ even argues that some of the employment at risk is in elite professional occupations, as technology changes the ways in which citizens access specialist expertise. And this in turn has led to speculation about the ‘hollowing-out’ of the middle class, with attendant worries about the future of democracy on the grounds that a stable middle-class is taken to be a necessary condition for liberal democracy.

Apocalyptic predictions about “robots taking our jobs” are nothing new — a fact that is currently used to dismiss fears about the technology. The reassuring lesson of history — so we are told — is that while machines have always displaced jobs, in the end the losses were compensated for by the emergence of new industries and new kinds of employment. The implicit assumption is that this history will repeat itself with AI. Perhaps it will. But the historical record also shows that those periods of transition were quite long and involved much hardship and disruption. One cautionary argument for taking seriously the potential employment threat from AI is that the pace of technical advance in digital technology means that societies may no longer have the time needed to adjust to the changed circumstances.

Wolfson College, Cambridge

6 September 2017

⁹⁹⁴ Chief Executive of the Allen Institute for Artificial Intelligence.

⁹⁹⁵ Park Benedikt Frey and Michael A. Osborne, “The Future of Employment: How Susceptible are Jobs to Automation?”, Oxford, Martin School, September 19, 2013 (http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf)

⁹⁹⁶ Richard Susskind and Daniel Susskind: *The Future of the Professions: How Technology Will Transform the Work of Human Experts*, OUP, 2016.

NCC Group plc – Written evidence (AIC0240)

1. Introductory comments

- 1.1. NCC Group is delighted to respond to the Artificial Intelligence Committee's invitation to provide our input to the Committee's ongoing considerations of artificial intelligence.
- 1.2. As a global cyber security and risk mitigation provider, we are acutely aware of and actively committed to understanding the cyber security implications of Artificial Intelligence (AI), Machine Learning (ML) and Automation. We are very pleased to offer an industry perspective.
- 1.3. We believe that much confusion has arisen around the terms Artificial Intelligence, Machine Learning and others. So as to aid clarity, in our discussions and work, and subsequent comments, we define these terms below and, where appropriate, use examples of what they practically mean in a cyber security context:
 - 1.3.1. **Artificial Intelligence (AI)** is an overarching term for systems that employ computer intelligence. This includes, for example, systems that can play games against humans, or those that can detect potential malicious behaviour within network traffic or audit logs.
 - 1.3.2. **Machine Learning (ML)**, for us, is a subfield of AI and computer science that provides computers with the ability to learn, without being explicitly programmed, when exposed to new data. This is done through the study and construction of algorithms that produce models from training data which are then used to make predictions on further data. In that context, *supervised learning* entails an algorithm being trained with labelled data, such as classification for malicious email detection; *unsupervised learning* entails the algorithm making its own decisions and inferences, such as the BotMiner system performing cluster analysis on network traffic; and *reinforcement learning* entails data being presented as a dynamic environment, such as in driverless car or AlphaGo technologies.
 - 1.3.3. A further subset of ML is **Deep Learning**, used primarily for computationally intensive AI functions, as part of which neural networks simulate the way in which the human brain processes information.
 - 1.3.4. Because ML typically requires training on large volumes of sample data, **Big Data** plays a key role. We understand Big Data as large volumes of data amassed by modern computer systems, which come with an inherent appetite for being mined in some way to derive value and insight from them. For example, network packet captures or audit

logs are processed in different ways to look for evidence of host or network intrusion.

- 1.4. Our internal Working Group, formed of a number of experienced ML practitioners and security consultants, is researching AI and ML from a number of different angles to understand the risks and opportunities of these technologies when applied to security problems, considering, amongst other things: “good” and “bad” applications of ML in cyber security; adversarial ML; learn taint; the impact of automated decision-making on GDPR and other regulations; and algorithm morality.
- 1.5. NCC Group does consider AI and ML as complementary tools to aid human cyber security efforts, but is ultimately interested in understanding to what extent AI and ML is hyperbole or emerging salvation. With that in mind, we look at developing assessment methodologies and strategies for security products and systems that employ attack-based AI and ML, as well as an understanding of the applications in our defensive services, provided by our Security Operations Centre (SOC) and Cyber Defence Operations teams.
- 1.6. In addition to our internal research, we are working closely with academia. This includes partnership collaborations with Royal Holloway University and University College London and a £500,000 research investment as part of CyberInvest.
- 1.7. We are delighted to see such in-depth parliamentary scrutiny of AI and ML technologies, both through the Artificial Intelligence Committee and the All-Party Group on Artificial Intelligence, and welcome recent Government announcements regarding AI and ML in cyber security, in the Industrial Strategy and the Interim Cyber Security Science & Technology Strategy. All these are welcome steps to ensure detailed consideration of this emerging technology and appropriate policy responses so as to turn the UK into a global centre of excellence.
- 1.8. That said, we would echo, too, the comments of others who have appeared in front of the Artificial Intelligence Committee. We believe it is crucially important:
 - 1.8.1. To maximise the advantages of both human and machine intelligence in cyber security applications, understanding AI and ML as complementary tools rather than replacements of human activities;
 - 1.8.2. To consider if legislative frameworks, such as the Computer Misuse Act 1990, remain fit for purpose or require updating to reflect new and emerging technologies and realities;
 - 1.8.3. To continue the fruitful collaboration between industry, academia and government, particularly where industry, such as ourselves, will have a significant role to play in helping to secure future AI and ML

systems, and critical infrastructure against the offensive use of AI and ML;

- 1.8.4. To avoid the development of silos and isolated initiatives across sectors. It is encouraging to see that cyber security is explicitly recognised as one of six priority business sectors as part of the AI Sector Deal in the Industrial Strategy, as well as identified as one of the emerging technologies that the Cyber Security Science and Technology Strategy sets out to address to ensure the UK stays ahead of the curve. We hope that the planned AI Council and Office for AI will allow for meaningful involvement of the cyber security industry, and continuous dialogue with the likes of the NCSC as research priorities and policy development is discussed;
- 1.8.5. To continue public and private investment in research and innovation, and skills development and expertise.

2. Detailed comments

- 2.1. Against that background, we have set out below our comments regarding the Artificial Intelligence Committee's areas of questioning. Our primary focus to date has been Machine Learning (ML), as a subfield of AI, and our comments should be read in that context:
- 2.2. Questions 1 – 2: What does artificial intelligence mean for cyber security today, and how is this likely to change over the next 10 years? Does artificial intelligence have implications for conventional cyber security today? Does AI facilitate new kinds of cyber-attacks, and if so, what are they? Are these potentially more dangerous or threatening? To what extent can AI help to strengthen cybersecurity? Where are such approaches used in cybersecurity, and how might this change in the future?
 - 2.2.1. We consider Machine Learning (ML), as a subfield of Artificial Intelligence (AI), to be a powerful tool, because it offers the possibility to retrain algorithms to adapt to different environments or changes in datasets. Indeed, the security industry has been quick to adopt the use of Machine Learning: it is well suited to the problem of classifying data from large data sets, and used across a range of products such as spam filtering, malware detection and network intrusion detection. ML-based applications such as natural language processing (NLP), moreover, offer potential value in open source intelligence (OSINT) analysis i.e. the identification and correlation of publicly available information and data from multiple sources, and other threat intelligence activities which require the reading, digesting and processing of (large) volumes of written text. OSINT analysis is an element of threat intelligence that

aims to assess (and minimise) the extent of information about an organisation that is freely available on the Internet. Malicious actors can leverage such exposure during their attacks e.g. to determine how a company operates, what its sources of profit are, and what potential entry points to its systems exist.

- 2.2.2. However, we do not consider ML as a panacea on its own. We strongly believe that the use of ML within security applications is most effective where objectives and desired outcomes are clearly defined from the outset, so that the most appropriate approach and algorithm can be determined. For example, identifying malware families on the basis of malware samples will require a different approach to classifying network logs into 'malicious' and 'safe' elements or making predictions using network traffic.
 - 2.2.3. In addition, NCC Group would caution against putting too much trust and too little human supervision in areas of AI and ML-led automation such as malware or breach detection. Specifically, we would express concerns with regard to the extent of "algorithmic authority" granted to AI and ML-based systems, where machines are making authoritative, yet critical decisions, which could have adverse effects on financials, safety, diplomacy etc.
- 2.3. Question 3: Will only state-sponsored hackers have the means to deploy AI in cyber-attacks? Or is there a risk that AI-enabled cyber-attacks will be democratised in the near future? Does this make a difference when attempting to defend against AI-enabled cyber-attacks? Are particular applications of AI, for example in healthcare or autonomous vehicles, more vulnerable to cyber-attacks than other areas, or is the threat quite evenly distributed across sectors?
- 2.3.1. We believe that it is inevitable that attackers will start using AI and ML for offensive operations. Tools are becoming more accessible, datasets are becoming bigger and skills are becoming more widespread, and once criminals decide that it is economically rational to use AI and ML in their attacks, they will.
 - 2.3.2. No particular field may be more vulnerable than another, however the potential infliction of physical harm in healthcare or autonomous vehicle applications may appeal to various hostile nation states, organised criminal or terrorist groups.
 - 2.3.3. The democratisation of technology availability risks inadvertent consequences too: as a result of ever-growing AI/ML frameworks becoming available to software developers that abstract data science and algorithmic details, developers will deploy ML and AI systems without necessarily understanding their underlying mathematics leading to potentially poor decisions.

2.4. **Question 4:** Do AI researchers need to be more aware of how their research might be misused, and consider how this might be mitigated before publishing? Are there situations where researchers should not publish or release AI research or applications with a high risk of misuse? Should the Government consider mechanisms, voluntary or mandatory, to restrict access in exceptional cases, in a similar way to the Defence Advisory Notice system for the media, for example?

2.4.1. We would argue that there are two core components to AI applications: (1) the AI algorithms themselves and (2) the data they use. In the end, any application is only as good and successful as the quality of data used to train it. As such, we contend that the publication of AI approaches and results need not be of any concern, unless the accompanying rich data sets are released alongside the algorithms.

2.4.2. In addition, we would draw to the Committee's attention two organisations that are actively considering how best to ensure safe and responsible AI research. OpenAI⁹⁹⁷ seeks to build "safe AI" while ensuring its benefits are "as widely and evenly distributed as possible". The Future of Life Institute⁹⁹⁸ has explored in detail the current state of discussions around the general "goal of keeping AI's impact on society beneficial".

2.5. **Question 5:** How much of an issue are recent developments in the field of adversarial AI for the wider deployment of AI systems? Should more attention be paid to adversarial AI attacks when developing new AI applications? Should mandatory regimes of stress-tested or penetration testing, prior to the release of systems or products, be required?

2.5.1. Note that our comments in this section are focused predominantly on ML-based systems and products, though, as per our definitions in the introductory comments, we consider ML to represent a subfield of AI.

2.5.2. Most ML-based products in use today are black box appliances that are placed onto networks and configured to consume data, process it and output decisions without humans having much knowledge of what's happening, giving adversaries a myriad of vectors available to attempt the manipulation of data that might ultimately affect operations. In addition, a growing number of online resources are available to support adversarial ML tasks.

2.5.3. ML algorithms are, by design, susceptible to influence and change; compromises and disruptions are usually achieved by manipulating data inputs to ML-based systems during the training, learning or operational

⁹⁹⁷ <https://openai.com/about/>

⁹⁹⁸ <https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/>

phase and seek to identify possible input that would be falsely classified, encourage misclassification, or adoption of “bad” behaviours. In that context, the most recent research in defence against adversarial ML focuses on detection and rejection of dangerous data before it reaches the classifier.

2.5.4. The black box nature of ML-based systems makes them hard to audit and assess using conventional testing methods. It is generally possible to perform static analysis but much more difficult to perform the more relevant dynamic analysis. By way of approaching audits more comprehensively, NCC Group has developed ten principles⁹⁹⁹ to consider as ML-based systems are implemented, covering: the datasets used for training the algorithm; its re-training frequency; adversarial testing and defence mechanisms; level of trust placed in decision-making, human supervision and override; and algorithmic transparency.

2.5.5. Moreover, it will be interesting to see if regulators continue to accept ML-based applications’ black box nature in security audits, particularly where such systems instigate key decisions such as financial transactions or security operations: the question of algorithmic transparency is likely to become increasingly important.

2.6. Question 6: How prepared is the UK for the impact of artificial intelligence on cyber security? Are the UK's national institutions sufficiently protected? Is the National Centre for Cyber Security doing enough? Should the National Cyber Security Strategy take explicit account of the threats and opportunities of AI for cyber security?

2.6.1. We welcome the recently published Interim Cyber Security Science & Technology Strategy which explicitly identifies AI, ML and automation as emerging technologies with implications for cyber security and sets out a framework to ensure that Government is adequately advised on the necessary policy responses to ensure the UK remains at the forefront of technology and has the capacity and capability to mitigate adversarial threats.

2.6.2. We believe that industry has an important role to play in those endeavours. We fear there is a risk that current political attention paid to AI technology might result in a plethora of separate initiatives that could fragment R&D and innovation efforts. We hope that there will be central coordination and streamlining of work underway so as to ensure industry is able best to navigate the landscape and contribute to maximum effect.

⁹⁹⁹ https://www.nccgroup.trust/globalassets/our-research/uk/whitepapers/2017/ncc_group_whitepaper_adversarial-machine-learning-approaches-and-defences.pdf (page 11)

2.6.3. NCC Group would also draw attention to the focused skills investment required to ensure genuine UK leadership in AI and ML which we would define as *producing* core AI frameworks, as opposed to *using* AI frameworks developed by others. Indeed, AI/ML are closely linked to data science and are very mathematical subjects. While there are many AI frameworks available for use that abstract away from the low-level minutiae/mathematics of AI, there is likely a major skills shortage of people with deep technical understanding of AI and its algorithms. There is therefore a danger that as a nation, the UK will be using AI frameworks developed by other nations, reliant on the assurances that they provide in the security of those frameworks. We strongly believe that this is a much less desirable outcome to being in a position where the UK is the producer of the core AI frameworks (that others might then use).

2.7. Question 7: Once the GDPR and the Data Protection Bill have come into force, will the law be able to adequately prosecute those who use misuse AI for criminal purposes?

2.7.1. We are generally concerned that a potentially outdated legislative framework inhibits the UK industry's ability to remain globally competitive in fields of emerging technologies related to cyber security, including AI and ML but also others such as threat intelligence and offensive cyber. For example:

2.7.1.1. Effective Machine Learning depends on the availability of large volumes of training data. Toy or public datasets are currently available for academic and industry researchers to use. Cyber-related data repositories and curated lists, covering anything from web attack payloads to malware samples, are often captured from online honeypots. We would urge the Committee to take into consideration any potentially necessary legislative reforms or regulatory updates to safeguard the continued availability and usability of such datasets and their procurement;

2.7.1.2. Depending on the application, it may be impossible to ascertain the source of any misuse of AI. If attackers can taint data used at training or operation phases, unless we are able to identify the source of any of those taints (which could be akin to finding a needle in a haystack), it might be extraordinarily difficult to prosecute criminals using traditional means.

2.7.2. We believe that considering the implications of emerging technologies for cyber security offers the opportunity more widely to review existing legislative and regulatory frameworks and ensure they

are fit for purpose to allow the UK to remain at the forefront of technological innovation and act as a global centre of excellence.

2.8. Question 8: How can personal data be effectively secured against misuse, especially given the potential conflict between secure and open data? Does the increasing availability of AI have implications for securing this data? How does artificial intelligence affect the security of anonymised datasets? Is there a level of anonymisation that is 'secure enough' to protect personal data against misuse? Are provisions in the Data Protection Bill sufficient to ensure that cyber-security researchers are able to test AI applications and data anonymisation protocols, without fear of legal prosecution? Is there a role for blockchain, and distributed-ledger technology more generally, in protecting personal data from AI-enabled cyberattacks?

2.8.1. As we outlined in our introductory comments, Big Data represents a major prerequisite for AI and ML. Simply put, AI is a consumer of data: the more data is available, and the more comprehensive it is, the better. This, of course, is largely antagonistic with principles of data protection and privacy.

2.8.2. We therefore believe that we will need to adopt a risk-based approach for each system, balancing the utility of an AI system with any compromise of user privacy. Consent, and consent revocation (like the right to be forgotten) will have to form a strong part of this (as per the provisions of the GDPR) though we do appreciate the complexities and difficulties involved in engineering any such consent methods into a complicated AI model. We therefore believe that it is imperative to ensure that implementers of AI carefully understand all of these nuances.

2.8.3. That said, anonymisation offers a good approach to tackle these complexities though does come with the shortfall of potentially reducing the utility of the data used to inform AI and ML models and algorithms.

2.8.4. Finally, we do not see any obvious role for blockchain/distributed ledger technologies in protecting personal data from AI-enabled cyberattacks.

2.9. Questions 9 – 10: How can we maintain the security of AI systems, particularly those of a safety-critical nature, both now and in the long term? Who should be responsible for securing and patching these systems, and how long should this responsibility be expected to last?

2.9.1. Principally, we believe that the key to maintaining AI systems' security will be data integrity and sanitisation. If the data used to train and operate AI systems can be tampered with, or tainted then it can render the system useless and/or untrustworthy from the outset. To

counter such risks, clear processes and mechanisms need to be in place by which AI applications carefully vet and sanitise their respective data supply chains, particularly where data originates from untrusted sources, such as the Internet and end-users.

2.9.2. In addition, we would advocate independent, third party product validation and additional research into required product updates for ML-based systems both to ensure their continued security, but also their long-term effectiveness as tools employed to safeguard other systems, networks and infrastructure.

2.9.3. In NCC Group's experience, current products often lack third party validation. Many claims made by ML product vendors, predominantly about products' effectiveness in detecting threats, are often unproven, or not verified by independent third parties.

2.9.4. In addition, more research is needed to understand the required updates for ML-based systems so as to ensure they don't become risks in themselves. As the cyber security industry has witnessed and learned with Intrusion Detection Systems (IDS), when detection signatures become outdated and are not maintained with signatures of the latest, emerging threats, the detection mechanism itself becomes less effective. The same issues may apply to ML-based systems whereupon their models/classifiers become outdated and ineffective. More needs to be done to understand the appropriate frequencies of model updates for different ML-based systems, and to understand better how best to perform dynamic model updating to determine what (human) supervision will be required for those processes.

2.10. Question 11: What is the one recommendation you would like to see this Committee make with its final report to the Government?

2.10.1. We would ask the Committee to describe AI and ML as a complementary part of cyber security, to be used as a tool or aide alongside rather than replacing human skill and scrutiny. AI and ML should not be seen as a panacea. Just like humans do, so AI and ML have their flaws; our endeavours should seek to maximise the advantages of both human and machine intelligence in cyber security applications rather than focus on one to the exclusion of the other.

2.10.2. In addition to this priority concern, we would ask the Committee to consider including in its final report the following recommendations:

2.10.2.1. A review of the current legislative framework to assess its fitness for purpose to accommodate emerging technologies and realities;

2.10.2.2. The continuation of collaboration between industry, academic and government, and continued commitment to public and

private investment in research, innovation and skills development; and

2.10.2.3. Strong coordination of existing government initiatives fostering AI, allowing for the meaningful involvement and contribution of the cyber security industry.

NCC Group plc

13 December 2017

Dr Jean-Christophe Nebel – Written evidence (AIC0102)

"The Select Committee on Artificial Intelligence was appointed by the House of Lords on 29 June 2017. It has been appointed to consider the economic, ethical and social implications of advances in artificial intelligence. It has to report by 31 March 2018".

My definition of artificial intelligence in the context of my response:

usage of a machine to make recommendations through automatic analysis of a large amount of data which is usually beyond what an expert could handle.

My answer addresses aspects of the following questions:

Impact on society

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved? In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

The process which has led to recommendations made by an artificial intelligence system can generally not be understood by a human, even an expert ('black box' effect). As a consequence, it is very difficult to challenge the output of such system. Moreover, it has been shown that, for example, a deep learning system can be relatively easily manipulated to produce any prediction which has serious security implications [Ng2015].

Artificial intelligence systems rely on large amount of data from which generalisations are made (system training) and used to make decisions. While such systems are proving to be more and more powerful and useful, one of their main drawbacks is their dealing with exceptional cases (outliers). By definition, an artificial intelligence system would not be trained for such cases (or trained insufficiently) and, as a consequence, would return uninformed decisions. There is a risk that users of artificial intelligence systems follow blindly their recommendations: first, systems are almost always right giving a false sense of

security; and, second, there may be a lack of awareness of the negative consequences resulting from incorrect decisions. Indeed, they may be either hidden by the much higher volume of successful outcomes or undetected due to outcomes taken place at a stage when causality with decision is difficult to establish. To prevent such situations, it is important that any result produced by such system is associated to not only appropriate confidence metrics, but also some clues about how a result has been obtained. They do not need to be comprehensive, but they need to be sufficient so that a user can challenge any obvious random or uninformed decision [Ng2015]. For example, key training examples which contributed to a recommendation could be highlighted or visualisation of 'neighbouring' cases associated with a similar outcome could be provided. In addition, a 'what if' functionality could be offered so that robustness of a decision could be tested by altering slightly the features of the case of interest. To summarise, humans need to be kept in the loop and novel tools should be developed so that they can interact with artificial intelligence systems and be able to exercise critical thinking even when faced with a black box.

Another important point is to be aware that artificial intelligence systems are NOT PC and do not have any agenda. Their task is to help decision making by generalising and, possibly, use stereotypes, if that leads to better global performance. If not moderated, such behaviour would be particularly unsuitable when dealing with decisions affecting directly vulnerable individuals.

Finally, artificial intelligence systems are firmly rooted in the past – they are based on past (training) examples – and, as a consequence, are unlikely to promote novel or creative solutions. This may lead to taking safe decisions with predictable outcomes, instead of novel or higher risk ones which could have much greater impact.

Reference

[Ng2015] A. Nguyen, J. Yosinski, J. Clune, "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images", in *Computer Vision and Pattern Recognition*, 2015

6 September 2017

Hadley Newman – Written evidence (AIC0155)

Artificial Intelligence in the United Kingdom

Submission to question 10: 'The role of the Government'

PRAGMATIC SOLUTIONS TO THE ISSUES PRESENTED BY ARTIFICIAL INTELLIGENCE:

WHAT ROLE SHOULD THE GOVERNMENT TAKE IN THE DEVELOPMENT AND USE OF ARTIFICIAL INTELLIGENCE IN THE UNITED KINGDOM? SHOULD ARTIFICIAL INTELLIGENCE BE REGULATED? IF SO, HOW?

Author: Hadley Newman, acting on an individual basis.

Hadley Newman is a Doctoral Researcher focused on the intersection of artificial intelligence, specifically human-machine communication, political marketing and voter behavioural intention, his Ph.D. is with the School of Social Sciences at Heriot-Watt University. He is Managing Director of a digital communications agency head quartered in Cambridge. Originally from London, Hadley has worked across key communication disciplines in Europe and MENA and is an Elected Fellow of The Royal Society of The Arts.

ARTIFICIAL INTELLIGENCE, DEFINED

Artificial intelligence (AI) refers to the science and engineering that enables computer systems to perform tasks that typically require human intelligence, including decision-making, speech recognition, visual perception, and communication. Subsets of AI include: (1) machine learning – i.e. programmed conditional statements and classification trees that enable machines to mimic natural intelligence, and (2) 'deep learning' which is best characterized as self-trained software (i.e. programmed algorithms) primed for recognition and communication. Much of AI's value is in human-machine communication, which has, effectively, blurred the boundaries between 'the artificial' and 'natural,' and plays a significant role in social development.

ABSTRACT

Artificial intelligence (AI) has advanced to the point where it can carry out complex human cognitive functions. Indeed, artificial neural networks (ANN) are continuously refined and closely replicate the structure and functions that used to be exclusive to human neural activity. This is not all to the good, however. There are real and perceived artificial intelligence hazards (AIH) to navigate, especially with respect to the socio-political and economic environments of nations worldwide, including the United Kingdom (UK). Specifically, political bots (cyber bots) are mainstays in "computational propaganda," and – as such – have added value to politicians and governmental actors and agents (i.e. cyber troops) who employ them. By that same token, nefarious users and groups have harnessed

bots to sway public opinion, stifle political critiques, miscast debates and sound-bites, and spread spam and propaganda throughout the UK. Responding to these threats and conflicts of interest, the government must position itself as an “independent mediator” on media platforms. In tandem, strong legislation and policy must be developed to guide the democratic infrastructure that facilitates AI discourse. In short, although AI’s use by political cyber troops remains in the best interests of national development and democracy, there must be checks and balances on its use by political stakeholders for the greater good of educating and informing the general public.

Keywords: Artificial Intelligence, Computational Propaganda, Political Bots, Political Cyber Troops.

Introduction and Background

1. Artificial intelligence (AI) refers to the science and engineering that enables computer systems to perform tasks that typically require human intelligence, including decision-making, speech recognition, visual perception, and communication (Chappell and Hawes, 2012). More precisely, artificial neural networks (ANN) are continuously refined and closely replicate the structure and functions that used to be exclusive to human neural activity (McFall and Mahan, 2009). With this functionality, AI has become a mainstay in socio-economic and industrial development (Hasler et al., 2013; Mou and Xu, 2017), seeing human-machine communications (HMC) and computer mediated communications (CMC) reach unprecedented levels (Mou and Xu, 2017). Still, there are real and perceived artificial intelligence hazards (AIH) to navigate on global, political, social, economic, and moral/ethical grounds. These threats range from calculated pre- and post-deployment threats (Yampolskiy, 2016) to imagined threats (Lapointe and Rivard, 2005; Weber Shandwick and KRC Research, 2016).
2. A recent UK study, carried out by Weber Shandwick and KRC Research (2016), revealed that the perceived threats of AI are higher in the UK than in other commonwealth nations. Of those surveyed, close to 90% of UK respondents expressed concerns about the use of AI for various criminal activities, including cyber-attacks and privacy issues, as well as loss of employment.
3. While AIH does pose a significant threat to the socio-political and economic environments of nations worldwide, including the UK, its outright ban would be detrimental on a large-scale – resulting in a loss of efficiencies and educational avenues (House of Commons Science and Technology Committee, 2016). An intermediary approach is, therefore, advisable – one that curbs negative social, political, legal, and ethical activities while promoting those that enhance social development and public interest (Bos-Nehles et al., 2017; Fang et al., 2014; House of Commons Science and Technology Committee, 2016; Siebert and Teizer, 2014).

Structure of Report

4. Organizationally, this report will begin by detailing the value added by AI and the challenges that come with it, specifically to the UK's political sphere. Solutions-oriented, this report will, ultimately, present a situated approach that the government can take to curb nefarious use.

AI's Impact on UK Society

5. Some well-described positive impacts of AI include creating efficiencies in UK industrial and economic systems (Fang et al., 2014; Leitão et al., 2016; Taylor, 2017). AI is particularly lauded in the UK's military (Bradshaw and Howard, 2017; Underwood, 2017) and energy sectors (Ahmad et al., 2014; Chou and Bui, 2014; Zaremba, 2017), which sees the use of drones in fire services, radiation leaks, and emergencies (Bos-Nehles et al., 2017; Shakmak and Al-Habaibeh, 2015; Swiss Foundation for Mine Action, 2016); and autonomous surface vehicle systems to check oil and gas (Ludvigsen and Sørensen, 2016; Siebert and Teizer, 2014).
6. By contrast, the downsides of AI have been framed as 'Artificial Intelligence Hazards' (AIH). Bostrom (2011) defines AIH as calculated computer-related risks, where the threat is derived from the sophistication of the cognitive functionality of the programs. A more nuanced description is offered by Yampolskiy (2016), who proposes a matrix of AIH contexts (internal-external) and timing (pre- and post-program deployment). So described, threats include artificial intelligence virus (AIV), spyware, Trojan horses, and intelligent worms (Yampolskiy, 2012).
7. It is worth noting that threats may be unintended, due to mistakes in program design (Dewey et al., 2015) or from recursive 'self-improvement,' 'self-delusions,' and 'self-wireheading' of AI (Nijholt, 2011; Yampolskiy, 2016).

Other threats (negative social/perceived)

8. As noted above, the threat of AI to employment and other industrial sectors was acknowledged by the UK government in a report released by the House of Commons' Committee for Science and Technology (2016). Consensus over a situated policy response, however, remains beyond reach.

Social interactions and computational propaganda

9. AI plays a key role in social media and public interactions, as well as freedom of speech and association (Mou and Xu, 2017) – which are fundamental human rights in the UK. With that, AI also has the power to change political discourse and ideologies (Hill et al., 2015; Howard and Kollanyi, 2016). The use of AI to influence the political sphere has been termed "computational propaganda" (CP) (Howard et al., 2016; Howard and Kollanyi, 2016; Kollanyi et al., 2016). Increasingly, concerns are expressed over the benefits and threats posed to the UK's democracy and government by CP in the form of 'political bots' (Howard and Kollanyi, 2016).

Political Cyber Troops vs. the Public

10. To delve deeper into the challenges presented by AI and propose pragmatic solutions in the political arena, I will present the interests of two major stakeholders. First are the political cyber troops who assume the forms of various actors, commissioning and managing political cyber AI as 'bots' (Bradshaw and Howard, 2017). Linked to this is the general public, who interact amongst themselves and with the political cyber troops and their cyber bots (Bradshaw and Howard, 2017). The central issues between these stakeholders are the undue influence of political ideologies, censorship, and the promotion of false information about political opponents.

Benefits to Political Troops

11. AI has proven beneficial to politicians worldwide (Woolley and Howard, 2016). Political cyber bots deliver political news, party updates, online feeds to party members and the general public, and other campaign messages using AI on social media platforms (Woolley, 2016). Political chatbots are used to interact with the public more intimately – e.g. they are often used to provide greater detail about a candidate's political platform (Bradshaw and Howard, 2017; Howard et al., 2016; Woolley, 2016).
12. Moreover, politicians and political parties benefit from the efficiency, cost, effectiveness, and versatility of AI (Dubbin, 2013; Yampolskiy, 2012). Essentially, AI streamlines processes and practices that would otherwise require employing multiple human agents. Political bots work around the clock, report feedback in real-time, and improve upon learned behaviour. They also cover large geographical spreads from a single location, reaching areas where it would be costly to send humans.

Benefits to the Public

13. Some of the benefits to the public include having a forum – with little to no forms of discrimination – to discuss issues such as renewable energy with their political representatives (Cass et al., 2010). Even the young, non-voting population of the UK have an avenue to state their concerns (Marsh et al., 2006).

Negative effects of AI in politics

14. Ultimately, it may be inferred from the discussion that there are limited negative repercussions to politicians themselves. After all, political cyber troops are the main actors who control the AI that informs political discourse. Hence, the negative implications are absorbed by the public and socio-political environment more broadly, insofar as AI communications are perceived as legitimate (Forelle et al., 2015; Howard, 2015). Research has shown that AI is often used to spread malicious content and political spam to achieve false political propaganda, often directed at political opponents' weak points and even their websites (Howard, 2015; Woolley, 2016; Howard and Kollanyi, 2016; Nazario, 2009).

15. The legal and moral implications of using AI in various sectors – not just politics – has yet to be properly legislated and documented (Weber Shandwick and KRC Research, 2016). Focusing on the UK political context, Bradshaw and Howard (2017) and Howard and Kollanyi (2016) observe AI's influence in the Brexit UK-EU Referendum, its outcome, and its cascading and enduring consequences.

The role of the Government as an "independent" mediator, free of conflict of interest

16. While a helpful example, the results of the UK-EU Referendum are not the focus of this discussion. Rather, this report highlights the conflicts between politicians and government political cyber troops. Either of these actors may work through private contractors (Monbiot, 2011), volunteers (Geybulla, 2016), or citizens (Kohli, 2013) to further their political interests. Considering the fact that politicians typically mature into the executive arm of government, governments need to assume an independent role in order to further public interest and maintain balanced political discourse.
17. It is encouraging that the UK's government has set up the 77 Brigade¹⁰⁰⁰ and The Government Communications Headquarters (GCHQ) to address conflicts of interest. The GCHQ has over 6000 employees and works in partnership with the Secret Intelligence Service (MI6) and MI5 to keep Britain safe and secure by examining AI's capabilities (GCHQ, 2017). Notably, the 77 Brigade - under the British Army - remains largely independent, despite the fact that the Army falls under the Monarch. Hence, political parties (even in government) require their own cyber troops to further their agendas.

The role of government in specific behaviours on media platforms

18. In addition to its independent role in maintaining balance in the UK's political discourse, the government may intervene to ensure that politicians do not abuse AI and dispel false information. To achieve this, information flow must be under the direct auspices of the government's AI policing. Because the greatest threat AI poses to democracy lies in the increasing meta-morph of social media platforms (Stern-Hoffman, 2013), the government can achieve balance through legislation. Specifically, the government must restrict the production of abusive and false information on the platform, while furthering its societally desirable agenda (Bradshaw and Howard, 2017).

The role of government in Legislation, Policies, and Democracy infrastructure

19. In legislating AI, it is essential to ensure that focus remains on the creation of better-informing citizens by educating them on their available choices. Any attempt to push the public in a specific direction would amount to a biased government that exerts undue influence. The fundamental rights of citizens include determining political outcomes of referendums and elections through

¹⁰⁰⁰ The 77 Brigade was announced in 2015 by the British Army to use non-lethal psychological tactics to combat insurgents, primarily through social media platforms of Facebook and Twitter.

votes. While politicians must be permitted to further their interests, the government must play a central role in propagating good media that results in educating and informing citizens. Outside of political discourse, a strong cyber security framework is critical to keep all AIH in check.

Conclusion

20. The report issued by the House of Commons' Science and Technology Committee on robotics and artificial intelligence (2016) acknowledges that nearly all areas of the UK public's life are affected by AI, and that AI is critical to the UK's global competitiveness. Employability and employment outcomes remain perceived threats by the public. Hence, despite projected positive contributions to productivity, collaborations with and upgrades in the labour industry and workforce are vital to the UK's labour-AI market. Ultimately, the UK Government must train and develop its labour force so as not to render their contributions obsolete and their roles expendable. Other threats can be handled through a strong cyber security legal infrastructure.
21. In conclusion, and with particular emphasis on the UK's political discourse, the use of AI by political cyber troops remains in the best interests of national development. AI is not only fundamental to the UK's socioeconomic environment but also to the public's well-being and democracy. Democracy requires that the public be educated and empowered to develop their own viewpoints and contribute to political discourse. That being said, politicians' use of AI should not be unbounded, as it may be detrimental to the UK's political society, as learned from Twitter's role in the 'Arab Spring' (Lotan et al., 2011) and the general influence of AI on global politics (Woolley, 2016). The government must maintain a neutral stance towards all players on the political platform, enabling the expression of political interests, ideologies, and platforms by political cyber troops in an effort to keep the public informed and educated. To accomplish this, the UK Government must remain proactive in future cyber trends in the political arena in order to cultivate a democratic yet competitive national environment.

REFERENCES

1. Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur, R., 2014. A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renew. Sustain. Energy Rev.* 33, 102–109.
2. Ai, W., 2012. China's paid trolls: Meet the 50-Cent Party. *New Statesman*. Retrieved from <http://www.newstatesman.com/politics/politics/2012/10/china%E2%80%99s-paid-trolls-meet-50-cent-party>

3. Bos-Nehles, A., Bondarouk, T., Nijenhuis, K., 2017. Innovative work behaviour in knowledge-intensive public sector organizations: the case of supervisors in the Netherlands fire services. *Int. J. Hum. Resour. Manag.* 28, 379–398.
4. Bostrom, N., 2011. Information hazards: a typology of potential harms from knowledge. *Rev. Contemp. Philos.* 10, 44.
5. Bradshaw, S., Howard, P.N., 2017. Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. *University of Oxford*. Retrieved from: <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/07/Troops-Trolls-and-Troublemakers.pdf>
6. Cass, N., Walker, G., Devine-Wright, P., 2010. Good Neighbours, Public Relations and Bribes: The Politics and Perceptions of Community Benefit Provision in Renewable Energy Development in the UK. *J. Environ. Policy Plan.* 12, 255–275. doi:10.1080/1523908X.2010.509558
7. Chappell, J., Hawes, N., 2012. Biological and artificial cognition: what can we learn about mechanisms by modelling physical cognition problems using artificial intelligence planning techniques? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367, 2723–2732. doi:10.1098/rstb.2012.0221
8. Chou, J.-S., Bui, D.-K., 2014. Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy Build.* 82, 437–446.
9. Dewey, D., Russell, S., Tegmark, M., et al., 2015. A Survey of Research Questions for Robust and Beneficial AI. *Future of Life Institute*. Retrieved from https://futureoflife.org/data/documents/research_survey.pdf
10. Dubbin, R., 2013. The Rise of Twitter Bots. *New Yorker*. Retrieved from <http://www.newyorker.com/tech/elements/the-rise-of-twitter-bots>
11. Fang, S., Da Xu, L., Zhu, Y., Ahati, J., Pei, H., Yan, J., Liu, Z., 2014. An integrated system for regional environmental monitoring and management based on internet of things. *IEEE Trans. Ind. Inform.* 10, 1596–1605.
12. Feldmann, H., 2013. Technological unemployment in industrial countries. *J. Evol. Econ.* 23, 1099–1126.
13. Filer, T., Fredheim, R., 2017. Popular with the Robots: Accusation and Automation in the Argentine Presidential Elections, 2015. *Int. J. Polit. Cult. Soc.* 30, 259–274.
14. Forelle, M.C., Howard, P.N., Monroy-Hernández, A., Savage, S., 2015. Political bots and the manipulation of public opinion in Venezuela. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2635800
15. GCHQ, 2017. Who we are | GCHQ. Retrieved from <https://www.gchq.gov.uk/who-we-are>

16. Geybulla, A., 2016. In the crosshairs of Azerbaijan's patriotic trolls. *ODR Russ. Beyond*. Retrieved from <https://www.opendemocracy.net/od-russia/arzu-geybulla/azerbaijan-patriotic-trolls>
17. Gibson, R.K., Lusoli, W., Ward, S., 2005. Online participation in the UK: Testing a 'contextualised' model of Internet effects. *Br. J. Polit. Int. Relat.* 7, 561–583.
18. Hasler, B.S., Tuchman, P., Friedman, D., 2013. Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Comput. Hum. Behav.* 29, 1608–1616.
19. Hill, J., Ford, W.R., Farreras, I.G., 2015. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Comput. Hum. Behav.* 49, 245–250.
20. House of Commons Science and Technology Committee, 2016. Robotics and artificial Intelligence. *House of Commons Science and Technology Committee*. Retrieved from <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>
21. Howard, P.N., 2015. *Pax Technica: How the Internet of things may set us free or lock us up*. New Haven, Connecticut: Yale University Press.
22. Howard, P.N., Kollanyi, B., 2016. Bots, #strongerin, and #brexit: Computational propaganda during the uk-eu referendum. (Comput. Propag. Proj. Research Note 2016.1). *Oxford Internet Institute, University of Oxford*. Retrieved from <https://arxiv.org/abs/1606.06356>
23. Howard, P.N., Kollanyi, B., Woolley, S.C., 2016. Bots and Automation over Twitter during the US Election. (Comput. Propag. Proj. Work. Pap. Ser. 2016.4 / 17). *Oxford Internet Institute, University of Oxford*. Retrieved from: <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2016/11/Data-Memo-US-Election.pdf>
24. Hughes, J.J., 2014. Are technological unemployment and a basic income guarantee inevitable or desirable. *J. Evol. Technol.* 24(1), 1-4
25. Kohli, K., 2013. Congress vs BJP: The curious case of trolls and politics. Retrieved from <http://timesofindia.indiatimes.com/india/Congress-vs-BJP-The-curious-case-of-trolls-and-politics/articleshow/23970818.cms>
26. Kollanyi, B., Howard, P.N., Woolley, S.C., 2016. Bots and automation over Twitter during the first US Presidential debate. (Comput. Propag. Proj. Data Memo 2016). *Oxford Internet Institute, University of Oxford*. Retrieved from <http://comprop.oii.ox.ac.uk/wp-content/uploads/2016/10/Data-Memo-First-Presidential-Debate.pdf>
27. Lapointe, L., Rivard, S., 2005. A multilevel model of resistance to information technology implementation. *MIS Q.* 461–491.

28. Leitão, P., Colombo, A.W., Karnouskos, S., 2016. Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges. *Comput. Ind.* 81, 11–25.
29. Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., others, 2011. The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *Int. J. Commun.* 5, 31.
30. Ludvigsen, M., Sørensen, A.J., 2016. Towards integrated autonomous underwater operations for ocean mapping and monitoring. *Annu. Rev. Control* 42, 145–157.
31. Marcus, G., 2014. Artificial Intelligence Isn't a Threat—Yet. *Wall Str. J.* 13–14.
32. Marsh, D., O'Toole, T., Jones, S., 2006. Young people and politics in the UK: Apathy or alienation? New York, N.Y: Springer Publishing
33. McFall, K.S., Mahan, J.R., 2009. Artificial neural network method for solution of boundary value problems with exact satisfaction of arbitrary boundary conditions. *IEEE Trans. Neural Netw.* 20, 1221–1233. doi:10.1109/TNN.2009.2020735
34. Mendonça, M., Angelico, B., Arruda, L.V.R., Neves, F., 2013. A dynamic fuzzy cognitive map applied to chemical process supervision. *Eng. Appl. Artif. Intell.* 26, 1199–1210. doi:10.1016/j.engappai.2012.11.007
35. Monbiot, G., 2011. The need to protect the internet from 'astroturfing' grows ever more urgent. *The Guardian*, 23. Retrieved from <https://www.theguardian.com/environment/georgemonbiot/2011/feb/23/need-to-protect-internet-from-astroturfing>
36. Mou, Y., Xu, K., 2017. The media inequality: Comparing the initial human-human and human-AI social interactions. *Comput. Hum. Behav.* 72, 432–440.
37. Nazario, J., 2009. Politically motivated denial of service attacks. *Virtual Battlef. Perspect. Cyber Warf.* 163–181.
38. Nijholt, A., 2011. No grice: computers that lie, deceive and conceal. *EEMCS*. Retrieved from <http://eprints.eemcs.utwente.nl/18432/>
39. Schroll, C., 2015. Splitting the Bill: Creating a National Car Insurance Fund to Pay for Accidents in Autonomous Vehicles. *Northwest. Univ. Law Rev.* 109, 803–833.
40. Shakmak, B., Al-Habaibeh, A., 2015. *Detection of water leakage in buried pipes using infrared technology; A comparative study of using high and low resolution infrared cameras for evaluating distant remote detection*. A paper presented at the Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference On. IEEE, pp. 1–7.

41. Siebert, S., Teizer, J., 2014. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. *Autom. Constr.* 41, 1–14.
42. SIHLOBO, W., 2017. Africa needs GM crops. *Farmer's Wkly.* 2017(17026), 13–23.
43. Stern-Hoffman, G.S., 2013. Government to use Citizens as Army in Social Media War. *Jerus. Post.* Retrieved from <http://www.jpost.com/Diplomacy-and-Politics/Government-to-use-citizens-as-army-in-social-media-war-322972>
44. Swiss Foundation for Mine Action, 2016. Using drones in fre and rescue services in the United Kingdom. *Swiss Foundation for Mine Action (FSD)*. Retrieved from <http://drones.fsd.ch/wp-content/uploads/2016/11/12.Manchester.pdf>
45. Taylor, I., 2017. Which UK Industries Will Benefit Most from Artificial Intelligence in 2017? CommsTrader. Retrieved from <https://www.commstrader.com/news/marketplace/uk-industries-will-benefit-artificial-intelligence-2017/>
46. Underwood, S., 2017. Potential and Peril: The outlook for artificial intelligence-based autonomous weapons. *Commun. ACM* 60, 17–19. doi:10.1145/3077231
47. Weber Shandwick and KRC Research, 2016. AI-Ready or Not: Artificial Intelligence here we come! *Weber Shandwick*. Retrieved from <http://www.webershandwick.com/uploads/news/files/AI-Ready-or-Not-report-Oct12-FINAL.pdf>
48. Woolley, S.C., 2016. Automating power: Social bot interference in global politics. *First Monday* 21. Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/6161>
49. Woolley, S.C., Howard, P.N., 2016. Automation, Algorithms, and Politics| Political Communication, Computational Propaganda, and Autonomous Agents—Introduction. *Int. J. Commun.* 10, 9.
50. Yampolskiy, R., 2012. Leakproofing the Singularity Artificial Intelligence Confinement Problem. *J. Conscious. Stud.* 19, 1–2.
51. Yampolskiy, R., Fox, J., 2012. Safety Engineering for Artificial General Intelligence. *Topio* 1–10.
52. Yampolskiy, R.V., 2016. *Taxonomy of Pathways to Dangerous Artificial Intelligence.*, In *AAAI Workshop: AI, Ethics, and Society*. (AAAI Technical Report WS-16-02). Palo Alto, California: The AAAI Press. Retrieved from <https://www.aaai.org/ocs/index.php/WS/AAAIW16/paper/viewFile/12566/12356>
53. Zaki, A., 2017. Autonomous vehicles: Impact on future stability of motor insurance. *Asia Insur. Rev.* Mar2017, 38–41.

54. Zaremba, H., 2017. Artificial Intelligence Is Crucial For The Energy Industry. *OilPrice.com* Retrieved from <http://oilprice.com/Energy/Energy-General/Artificial-Intelligence-Is-Crucial-For-The-Energy-Industry.html>

6 September 2017

Hadley Newman – Supplementary written evidence (AIC0156)

Artificial Intelligence in the United Kingdom

Submission to question 11: 'Learning from others'

PRAGMATIC SOLUTIONS TO THE ISSUES PRESENTED BY ARTIFICIAL INTELLIGENCE:

WHAT LESSONS CAN BE LEARNT FROM OTHER COUNTRIES OR INTERNATIONAL ORGANISATIONS IN THEIR POLICY APPROACH TO ARTIFICIAL INTELLIGENCE?

Author: Hadley Newman, acting on an individual basis.

Hadley Newman is a Doctoral Researcher focused on the intersection of artificial intelligence, specifically human-machine communication, political marketing and voter behavioural intention, his Ph.D. is with the School of Social Sciences at Heriot-Watt University. He is Managing Director of a digital communications agency head quartered in Cambridge. Originally from London, Hadley has worked across key communication disciplines in Europe and MENA and is an Elected Fellow of The Royal Society of The Arts.

ARTIFICIAL INTELLIGENCE, DEFINED

Artificial intelligence (AI) refers to the science and engineering that enables computer systems to perform tasks that typically require human intelligence, including decision-making, speech recognition, visual perception, and communication. Subsets of AI include: (1) machine learning – i.e. programmed conditional statements and classification trees that enable machines to mimic natural intelligence, and (2) 'deep learning' which is best characterized as self-trained software (i.e. programmed algorithms) primed for recognition and communication. Much of AI's value is in human-machine communication, which has, effectively, blurred the boundaries between 'the artificial' and 'natural,' and plays a significant role in social development.

ABSTRACT

Increasingly, artificial intelligences (AI) can adapt to their changing environments, 'thinking' and 'behaving' like humans. While this move towards independence and autonomy signals advancement and creates efficiencies, there are real and perceived artificial intelligence hazards (AIH) to navigate. This is especially the case in socio-political and economic environments of world nations, including the United Kingdom (UK). Yet, the realm of AI policy remains broad, ambiguous, and speculative. At present, established laws and regulatory bodies

are stretched to include emergent AI issues in the UK. A resilient and tailored policy approach requires implementing the European Union General Data Protection Regulation (GDPR) to bolster the original Data Protection Act passed in 1998. Essentially, the UK needs to add to the GDPR to anticipate critical issues pertaining to the autonomous behaviour of AI. In the domain of cyber regulation, the Computer Misuse Act (CMA) is equally outdated and requires significant improvements. As internet and computer applications increase in scope, access, and capabilities, cybercrime laws must take into consideration the delegated or distributed authority of autonomous programs by learning from the Human Rights Watch (HRW) and the UK National Crime Agency. The UK can also improve upon cybercrime policies by drawing on the best practices of other countries. Finally, strict laws must be adopted to oversee social media behaviour. In short, a comprehensive, 'whole of government' approach must be taken to ensure effective and efficient AI policy implementation in the UK, building on key insights gathered from the Australian Public Service Committee.

Keywords: Artificial intelligence, Human Machine Communication, Data Protection, Cybercrime, Social Media, Whole of Government

Introduction and Background

1. Artificial intelligence (AI) refers to the science and engineering that enables computer systems to perform tasks that typically require human intelligence, including decision-making, speech recognition, visual perception, and communication (Chappell and Hawes, 2012). More precisely, artificial neural networks (ANN) are continuously refined and closely replicate the structure and functions that used to be exclusive to human neural activity (McFall and Mahan, 2009). Of particular interest is the role of AI in human-machine communications (HMC) (Howard and Kollanyi, 2016; Mou and Xu, 2017).
2. The European Parliament, United States' White House, and the United Kingdom's House of Common's reports all appreciate the importance of preparing citizens for the widespread use of AI (Cath et al., 2017). Specifically, they note the unique advantages of HMCs - as ongoing sense-making processes between humans and machines (Mou and Xu, 2017) - and recognise AIs as social agents.

HMCs, Social Media and AI in Politics

3. HMCs offer a cheap, efficient, and effective alternative to human resource capabilities in engaging audiences on social media (Howard and Kollanyi, 2016). Facilitated by AI, the role of social media in public engagement has grown exponentially in recent years (Senadheera et al., 2015) to the point where it has become a credible communications medium (Kane et al., 2012).
4. Correspondingly, ethical questions have been raised about the use of chatbots to further political agendas (Howard and Kollanyi, 2016). To date, AI continues to play a central role in global politics, evidenced in the 2010 Australian elections

(Bruns and Burgess, 2011), Swedish elections (Larsson and Moe, 2012), the UK-EU Referendum (Howard and Kollanyi, 2016), and the 2016 American presidential election (Kollanyi et al., 2016). Political campaigns today use hundreds - even thousands - of chatbots, publishing over a million Tweets daily (Gunnarsson Lorentzen, 2014; Howard and Kollanyi, 2016).

Problem and Structure of Report

5. According to Bex et al. (2017) "AI is currently at the centre of [the] attention of legal professionals." There is equal focus on how issues arising from the application of AI can be legally resolved, and how AI can be applied to enhance efficiency and effectiveness in the law's implementation.
6. Notably, there is consensus over the lack of strict guidelines to govern AI law, policy, and ethics in the UK (Ashrafian, 2015; Earl, 2016). Few to no physical laws exist with regard to the physical limits or confines of AI modelling and application (Hawking et al., 2014). In 2016 alone, the UK lost over £1billion to cybercrime (Wilson, 2017). The unaccounted effect on jobs and overall employment in the UK, based on perceived impact, is over 91% (Weber Shandwick and KRC Research, 2016).
7. More threatening, the program-based hazards of AI have been observed by Bostrom (2011) and Yampolskiy (2016). These harms result from direct interaction with AI programs. According to Smith (2015), AI in and of itself is not a threat; rather the real threat stems from "... our failure to create a policy framework for emerging technology (p. 1)."
8. Highlighting the need to remain proactive on policies surrounding emerging technology, the present report commences with an introduction to AI policy laws and ethics in the UK. Key policy areas - such as those regarding data protection, policies on UK cyber activities, social media regulations, and finally a holistic AI policy implementation approach - will be presented.

Introduction to the realm of AI policy, law, and ethics

9. AI policy is broad, ambiguous, and speculative (Chessen, 2017; Earl, 2016). In an effort to cover the broad scope of AI application (Chessen, 2017), policies range from ethics, transparency, norms, trust, information security, economics, and privacy to humanitarian issues. Other policies and regulatory concerns have been raised in the realm of information access, malicious use of AI, cyber security, and autonomous weapons systems (Chessen, 2017).
10. Considering there is no clear and exhaustive framework to guide policymaking in this area, the present report draws on three main prongs of AI policy formation to propose a holistic approach to AI policy implementation in the UK. To that end, existing policies applicable to these areas are first discussed, followed by key improvement avenues that may be adapted by learning from other countries and global organisations. This is essential to provide the necessary legal

framework needed to guide the application of AI to various areas of public life, while, at the same time, curbing its associated challenges in the UK.

Improvement in the UK Data Protection Act's Policies

11. The UK Data Protection Act was formulated in 1998 and came into effect in 2000, replacing the EU Data Protection Directive of 1995 (UK Government, 1998). Even though some amendments have been made over the years (UK Government, 2017), the Act still requires a significant overhaul (Chessen, 2017). The need to improve upon data protection laws to accommodate the growth in AI application has been acknowledged across Europe (Wilson, 2017). A prime candidate to model is GDPR, which is earmarked for implementation throughout Europe (projected for 2018). For the UK's purposes, this framework can be converted into British Law through the Great Repeal Act.
12. The framework may still require additional components, however. Key considerations in the recent amendments to the Data Protection Act and the GDPR include the acknowledgement of data protection issues surrounding the processing and movement of data, even in domestic use. This inclusion is critical, considering a significant portion of the original UK Data Protection Act does not apply domestically (UK Government, 1998). Information and data currently lie within public reach, not merely within institutions. A critical analysis of the GDPR also reveals flaws in handling the full scope of data automation areas (Fulford and Lockyer, 2016).
13. Given these considerations, there is a need to take into full consideration the autonomous behaviour of AI, and the autonomous collection, management, and distribution of data by intelligent programs operating without human interference (Fulford and Lockyer, 2016). Adding to these improvements, concerns about data protection must cover areas of fairness, transparency, accuracy, security, accountability, data controllers, and other implications of AI to data protection, as articulated by the Information Commissioner's Office (Information Commissioner's Office (ICO), 2017).

Policies on UK Cybercrime and Cyber Regulations

14. The Internet of Things presents challenges for data protection (Cooper and James, 2009; Cumbley and Church, 2013). Most devices that are near humans will soon have a dedicated IP address and internet connection capabilities. This will significantly increase machine-to-machine communication as well as human-machine-communication (Cooper and James, 2009). By that same token, this phenomenon will increase the scope, access, and capabilities of cyber criminals. It is, therefore, critically important that cyber laws take effect and bear on all areas of "Big Data" exploitation (Cumbley and Church, 2013). Specifically, cybercrime laws must take into account the delegated or distributed authority of autonomous programs (Adam, 2005), given that these areas are barely touched on in the UK Computer Misuse Act (CMA) of 1990. In fact, the CMA lacks even a standing definition of the terms "computer" and "AI". Even with the support of

the Regulation of Investigatory Powers Act (2000) and the Data Protection Act (1998), serious flaws remain that bear on the capabilities of the CMA to tackle cybercrime in the UK today (The Crown Prosecution Service, 2017a).

Strengthening the case, the Office for National Statistics (ONS) found over 5.6 million fraud and computer misuse crimes for the year-end June 2016 (Scott, 2016). This cost the UK over £1billion (Wilson, 2017). Remarkably, the CMA considers cybercrime in only two areas, i.e. cyber-dependent and cyber-enabled crimes.

15. The UK can improve upon this Act by drawing on learnings and best practices from other countries that have considered automated cybercrimes. As defined by Parker (1995), cybercrimes have no human interference from start to finish. These crimes are committed based on the self-delusions and recursive self-improvement abilities of AI (Nijholt, 2011; Yampolskiy, 2016). The question of whether programmers, manufacturers, and military personnel can be held liable for the activities of fully automated machines - like cars and surveillance robots - has been raised by Human Rights Watch (HRW) (Green, 2015). Experts anticipate that computer-automated crimes will increase exponentially by 2040 (Duncan, 2016). The expansion of the cybercrime law to cover all of these areas is therefore critical, as noted by the National Crime Agency (2016).
16. The transnational nature of cybercrimes must also be accounted for in any thoroughgoing policy response. Indeed, more cyber and computer crimes take place across national boundaries (Huang and Wang, 2009), making “cyber war” a daunting reality. For instance, the recent allegations of Russia meddling in the USA’s 2016 presidential election have drawn tensions similar to those of the cold war era (Kreps and Das, 2017; Ohlin, 2016). Hence, the EU approach to Cyber Security must take into consideration crimes beyond national borders and international protections (European Parliamentary Research Service, 2014). The UK can further tap into global surveillance systems, as recommended by the United Nation’s Human Rights Council (Miles, 2017). In fact, Cuba and Iran – countries with strict media – embraced this treaty.

Policies on Social Media Restrictions

17. A number of European countries have taken bold steps to pass social media restrictions into law (Marche, 2012; Turkle, 2012). Several states in the USA - including Washington, New Jersey, and California - have strict legislative controls on employers, prohibiting them from requesting the credentials of their employees, and from employees being punished by employers based on their social media activities. Other states have focused on students and other post-secondary institutions, in addition to the laws on employer-employee social media relationships (Turkle, 2012).
18. In the UK, the government may be applauded for establishing guidelines on the legal approach to cases involving communication via social media (The Crown Office and Prosecutor Fiscal Service, 2017; The Crown Prosecution Service, 2017b). Even though these guidelines cover a wide range of areas, from hate

crimes to social media restraining orders, it must be strictly enforced or passed as an Act to ensure strict adherence. Currently, most legal cases on social media are handled under different sets of policies and Acts that are retroactively applied (Earl, 2016).

19. Aside from the need to pass a comprehensive Act on social media that restricts data sharing between actors in the workplace, the UK can, again, learn from Asian countries, where strict social media regulations are enforced without having to maintain an autocratic government that disregards fundamental human rights, like those of North Korea and Syria. The social media policies in Japan, Singapore, and Israel are instructional (Bradshaw and Howard, 2017; Stern-Hoffman, 2013). These countries' measures entail strict implementation of the authority of command in the workplace, and prevention of the spread of unethical, immoral, and false news (Stern-Hoffman, 2013).
20. The UK can also observe the practices of their European counterparts with respect to social media surveillance. For example, the influx of refugees and prevalence of terrorism in France, Germany, and Spain have motivated these countries to oversee public data with little exception (Nossel, 2016). In fact, the state of emergencies enforced during the recent terrorist attacks in France granted the Government the power to control the press, media, videos, plays, and all forms of personal data throughout the country, over the extended period of the emergency.
21. Instead of putting such systems in place after emergencies, a constant surveillance and monitoring policy should be implemented to at least sift through some depth of all data forms in the UK. Supporting this practice, advanced AI algorithms can be implemented to automatically retrieve only the data that might pose a threat to the UK and its allies.

Whole of Government Approach to Policy Implementation

22. The need to incorporate a Whole of Government Approach (WGA) to policy implementation is presented to provide an all-inclusive and unified approach to dealing with issues arising from AI. This approach to policy implementation entails that all public services and agencies work together in an integrated governmental response to relevant issues (Christensen and Laegreid, 2007; Commission, 2012; Australian Public Service Committee, 2015). More precisely, in Australia, a cabinet implementation unit was set up for the express purpose of compelling sectorial authorities to coordinate and cooperate in tackling problems brought about by AI implementation.
23. It is important to emphasise that even though the WGA has been implemented in conventional areas of public life in several countries - including Canada, New Zealand, Singapore, and the USA (Aucoin, 2006; Boston and Eichbaum, 2005) - its potential use in controlling issues arising from AI is only beginning to be recognised. A number of reports and national plans by various public institutions

in Australia, such as the Whole of Nation in Cyberpower (Klimburg, 2015) and the National Plan to Combat Cybercrime (Commonwealth of Australia, 2013), take into keen consideration the role of inter-departmental and inter-agency cooperation. In the USA, Carlin (2015) claims a WGA is imperative to detect, disrupt, and deter national security cyber threats.

Conclusion

24. In conclusion, the supportive role that AI plays in the UK's public life is generally acknowledged (House of Commons Science and Technology Committee, 2016). However, threats have been registered, and the cost they present to the global economy is severe. Accordingly, careful regulation through policy formulation and a holistic implementation approach is paramount to effectively combat and reduce the threats that will, inevitably, arise as the result of advanced cyber applications. The UK Government must remain proactive in anticipating future cyber trends in order to put in place and maintain a resilient policy approach.

REFERENCES

1. Adam, A., 2005. Delegating and distributing morality: Can we inscribe privacy protection in a machine? *Ethics and Information Technology* 7, 233–242. Retrieved from <http://www.springerlink.com/index/v41x443r1t470231.pdf>.
2. Ashrafian, H., 2015. AIonAI: A humanitarian law of artificial intelligence and robotics. *Science and engineering ethics* 21, 29–40. Retrieved from <http://link.springer.com/article/10.1007/s11948-013-9513-9>.
3. Aucoin, P., 2006. Accountability and coordination with independent foundations: A Canadian case of autonomization. *Autonomy and regulation: Coping with agencies in the modern state* 110–36.
4. Australian Public Service Committee, 2015. Connecting Government: Whole of government responses to Australia's priority challenges. *Australian Government*. Retrieved from <http://www.apsc.gov.au/publications-and-media/archive/publications-archive/connecting-government>.
5. Bex, F., Prakken, H., van Engers, T., Verheij, B., 2017. Introduction to the special issue on Artificial Intelligence for Justice (AI4J). *Artificial Intelligence and Law* 25, 1–3. Retrieved from <http://link.springer.com/article/10.1007/s10506-017-9198-5>.
6. Boston, J., Eichbaum, C., 2005. State sector reform and renewal in New Zealand: lessons for governance, in: Conference on 'Repositioning of Public Governance—Global Experiences and Challenges', *Taipei*. pp. 18–19.

7. Bostrom, N., 2011. Information hazards: a typology of potential harms from knowledge. *Review of Contemporary Philosophy* 10, 44. Retrieved from <http://www.cceol.com/content-files/document-42884.pdf>.
8. Bradshaw, S., Howard, P.N., 2017. Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation. *Computational Propaganda Research Project*. Retrieved from <http://www.eluniverso.com/sites/default/files/archivos/2017/07/troops-trolls-and-troublemakers.pdf>.
9. Bruns, A., & Burgess, J., 2011. # ausvotes: How Twitter covered the 2010 Australian federal election. *Communication, Politics & Culture* 44, 37. Retrieved from <https://search.informit.com.au/documentSummary;dn=627330171744964;res=IELAPA>.
10. Carlin, J.P., 2015. Detect, Disrupt, Deter: A Whole-of-Government Approach to National Security Cyber Threats. *Harv. Nat'l Sec. J.* 7, 391. Retrieved from http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/harvardnsj7§ion=9.
11. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., Floridi, L., 2017. Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. *Science and Engineering Ethics* 1–24. Retrieved from <http://link.springer.com/article/10.1007/s11948-017-9901-7>.
12. Chappell, J., Hawes, N., 2012. Biological and artificial cognition: what can we learn about mechanisms by modelling physical cognition problems using artificial intelligence planning techniques? *Philosophical Transactions of The Royal Society Of London. Series B, Biological Sciences* 367, 2723–2732. Retrieved from . doi:10.1098/rstb.2012.0221
13. Chessen, M., 2017. The AI Policy Landscape. Artificial intelligence policy, laws and ethics. *Medium Inc.* Retrieved from <https://medium.com/artificial-intelligence-policy-laws-and-ethics/the-ai-landscape-ea8a8b3c3d5d>.
14. Christensen, T., L'Arronde, P., 2007. The whole-of-government approach to public sector reform. *Public administration review* 67, 1059–1066. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6210.2007.00797.x/full>.
15. Clarke, R.A., Knake, R.K., 2014. Cyber war. *Tantor Media, Incorporated*. Retrieved from [http://indianstrategicknowledgeonline.com/web/Cyber War The Next Threat to National Security and What to Do About It \(Richard A Clarke\).pdf](http://indianstrategicknowledgeonline.com/web/Cyber War The Next Threat to National Security and What to Do About It (Richard A Clarke).pdf).
16. Commission, A.P.S., 2012. Tackling wicked problems: A public policy perspective. *APSC*. Retrieved from <http://www.apsc.gov.au/publications-and-media/archive/publications-archive/tackling-wicked-problems>.

17. Commonwealth of Australia, 2013. National Plan to Combat Cybercrime. *Business and Information Law Branch, Attorney-General's Department*. Retrieved from [https://www.ag.gov.au/CrimeAndCorruption/Cybercrime/Documents/National Plan to Combat Cybercrime.pdf](https://www.ag.gov.au/CrimeAndCorruption/Cybercrime/Documents/NationalPlan%20to%20Combat%20Cybercrime.pdf).
18. Cooper, J., James, A., 2009. Challenges for database management in the internet of things. *IETE Technical Review* 26, 320–329. Retrieved from <http://www.tandfonline.com/doi/abs/10.4103/0256-4602.55275>.
19. Cumbly, R., Church, P., 2013. Is "big data" creepy? *Computer Law & Security Review*, 29, 601–609. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0267364913001349>.
20. Duncan, J., 2016. Robots and computers will commit more crime than humans by 2040, expert warns. MAILONLINE. Retrieved from <http://www.dailymail.co.uk/news/article-3780314/Robots-computers-commit-crime-humans-2040-expert-warns.html>.
21. Earl, A., 2016. The Law and Social Media. *UK Safer Internet Centre*. Retrieved from <https://www.saferinternet.org.uk/blog/law-and-social-media>.
22. European Parliamentary Research Service, 2014. EU approach to cyber-security. *European Parliamentary Research Service*. Retrieved from [http://www.europarl.europa.eu/RegData/bibliotheque/briefing/2014/140775/LDM_BRI\(2014\)140775_REV1_EN.pdf](http://www.europarl.europa.eu/RegData/bibliotheque/briefing/2014/140775/LDM_BRI(2014)140775_REV1_EN.pdf).
23. Fulford, N., Lockyer, G., 2016. Data Protection in the new world of Artificial Intelligence (UK Report No. Issue 85). *Privacy Laws & Business*. Retrieved from http://www.kemplittle.com/cms/document/Data_Protection_in_the_New_World_of_Artificial_intelligence.pdf.
24. Green, C., 2015. Computer programmers, manufacturers and military personnel would all escape liability for unlawful deaths caused by fully autonomous weapons. *Independent*. Retrieved from <http://www.independent.co.uk/life-style/gadgets-and-tech/news/killer-robots-no-one-liable-if-future-machines-decide-to-kill-says-human-rights-watch-10165653.html>.
25. Gunnarsson Lorentzen, D., 2014. Polarisation in political Twitter conversations. *Aslib Journal of Information Management* 66, 329–341. Retrieved from <http://www.emeraldinsight.com/doi/abs/10.1108/AJIM-09-2013-0086>.
26. Hawking, S., Russell, S., Tegmark, M., Wilczek, F., 2014. Stephen Hawking: Transcendence looks at the implications of artificial intelligence-but are we taking AI seriously enough? *The Independent*. Retrieved from <http://www.citeulike.org/group/15400/article/13524721>.

27. House of Commons Science and Technology Committee, 2016. Robotics and artificial Intelligence. *House of Commons Science and Technology Committee*. Retrieved from <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>.
28. Howard, P.N., Kollanyi, B., 2016. Bots, # strongerin, and # brexit: Computational propaganda during the UK-EU referendum. *Comprop Data Report*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2798311.
29. Huang, W., Wang, S.-Y.K., 2009. Emerging cybercrime variants in the Socio-Technical space, in: *Handbook of Research on Socio-Technical Design and Social Networking Systems*. IGI Global, pp. 195–208.
30. Information Commissioner’s Office (ICO), 2017. Big data, artificial intelligence, machine learning and data protection (Data Protection Act and General Data Protection Regulation No. Version: 2.1). *Information Commissioner’s Office*. Retrieved from <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.
31. Kane, G.C., Alavi, M., Labianca, G.J., Borgatti, S., 2012. What’s different about social media networks? A framework and research agenda. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2239249.
32. Klimburg, A., 2015. The Whole of Nation in Cyberpower. *Austrian Institute for International Affairs*. Retrieved from http://journal.georgetown.edu/wp-content/uploads/2015/07/171_gj124_Klimburg-CYBER-2011.pdf.
33. Kollanyi, B., Howard, P.N., Woolley, S.C., 2016. Bots and automation over Twitter during the first US Presidential debate. *Comprop Data Memo*. Retrieved from <https://assets.documentcloud.org/documents/3144967/Trump-Clinton-Bots-Data.pdf>.
34. Kreps, S., Das, D., 2017. Warring from the virtual to the real: Assessing the public’s threshold for war over cyber security. *Research & Politics* 4, 2053168017715930. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/2053168017715930>.
35. Larsson, A.O., Moe, H., 2012. Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society* 14, 729–747. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/1461444811422894>.
36. Madsen, W., 1992. *Handbook of personal data protection*. Springer. Retrieved from <https://www.google.com/books?hl=en&lr=&id=BruxCwAAQBAJ&oi=fnd&pg=PR1>

1&dq=UK+Data+Protection+Act+and+artificial+intelligence&ots=t8bWaiWyG9&sig=Dt5S3LE7TOkNW-vbDXObTBN0w94.

37. Marche, S., 2012. Is Facebook Making Us Lonely? *The Atlantic*. Retrieved from <https://www.theatlantic.com/magazine/archive/2012/05/is-facebook-making-us-lonely/308930/>.
38. McFall, K.S., Mahan, J.R., 2009. Artificial neural network method for solution of boundary value problems with exact satisfaction of arbitrary boundary conditions. *IEEE Transactions on Neural Networks* 20, 1221–1233. Retrieved from. doi:10.1109/TNN.2009.2020735
39. Miles, T., 2017. U.N. expert urges states to work toward cyber surveillance treaty. *Reuters*. Retrieved from <https://www.reuters.com/article/us-privacy-un/u-n-expert-urges-states-to-work-towards-cyber-surveillance-treaty-idUSKBN16F1M3>.
40. Mou, Y., Xu, K., 2017. The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, 72, 432–440. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0747563217301486>.
41. National Crime Agency, 2016. NCA Strategic Cyber Industry Group Cyber Crime Assessment 2016. *National Crime Agency*. Retrieved from <http://www.nationalcrimeagency.gov.uk/publications/709-cyber-crime-assessment-2016/file>.
42. Nijholt, A., 2011. No grice: computers that lie, deceive and conceal. *EEMCS*. Retrieved from <http://eprints.eemcs.utwente.nl/18432/>.
43. Nossel, S., 2016. Europe’s Free-Speech Apocalypse Is Already Here. *Foreign Policy*. Retrieved from <https://foreignpolicy.com/2016/03/17/europes-free-speech-apocalypse-is-already-here-france-germany-spain/>.
44. Ohlin, J.D., 2016. Did Russian Cyber Interference in the 2016 Election Violate International Law. *Tex. L. Rev.* 95, 1579. Retrieved from http://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/tlr95§ion=58.
45. Parker, D.B., 1995. Defining Automated Crime. *Information Systems Security* 4, 16–21. Retrieved from <http://dx.doi.org/10.1080/10658989509342504>. doi:10.1080/10658989509342504
46. Scott, P., 2016. How much of a problem is cyber-crime in the UK? *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/news/2016/11/01/how-much-of-a-problem-is-cyber-crime-in-the-uk/>.

47. Senadheera, V., Warren, M.J., Leitch, S., Pye, G., 2015. *Adoption of Social Media as a Communication Medium: A Study of Theoretical Foundations.*, In: UKAIS. p. 10. Retrieved from <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1009&context=ukais2015>.
48. Smith, A., 2015. Artificial intelligence. National Magazine. *National Magazine*. Retrieved from <http://www.nationalmagazine.ca/Articles/Fall-Issue-2015/Artificial-intelligence.aspx>.
49. Stern-Hoffman, G.S., 2013. Government to use Citizens as Army in Social Media War. *The Jerusalem Post*. Retrieved from <http://www.jpost.com/Diplomacy-and-Politics/Government-to-use-citizens-as-army-in-social-media-war-322972>.
50. The Crown Office and Prosecutor Fiscal Service, 2017. Guidance on cases involving Communications sent via Social Media. *The Crown Office and Prosecutor Fiscal Service*. Retrieved from http://www.copfs.gov.uk/images/Documents/Prosecution_Policy_Guidance/Book_of_Regulations/Final%20version%2026%2011%2014.pdf.
51. The Crown Prosecution Service, 2017a. Cybercrime - Legal Guidance. *The Crown Prosecution Service*. Retrieved from http://www.cps.gov.uk/legal/a_to_c/cybercrime/.
52. The Crown Prosecution Service, 2017b. Guidelines on prosecuting case involving communications sent via social media. *The Crown Prosecution Service*. Retrieved from http://www.cps.gov.uk/legal/a_to_c/communications_sent_via_social_media/.
53. Turkle, S., 2012. *Alone together: Why we expect more from technology and less from each other*. Basic Books, New York, N.Y.
54. UK Government, 2017. UK Data Protection Act 1998. *UK Government*. Retrieved from <http://www.legislation.gov.uk/ukpga/1998/29/introduction>.
55. UK Government, 1998. UK Data Protection Act 1998. *UK Government* Retrieved from <http://www.legislation.gov.uk/ukpga/1998/29/introduction/enacted>.
56. Weber Shandwick and KRC Research, 2016. AI-Ready or Not: Artificial Intelligence Here We Come! *Weber Shandwick* Retrieved from <http://www.webershandwick.com/uploads/news/files/AI-Ready-or-Not-report-Oct12-FINAL.pdf>.
57. Wilson, H., 2017. European data protection laws are changing. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/connect/small-business/business-networks/bt/data-protection-laws-changing/>.

58. Yampolskiy, R.V., 2016. *Taxonomy of Pathways to Dangerous Artificial Intelligence.*, in: *AAAI Workshop: AI, Ethics, and Society*. Retrieved from <http://www.aaai.org/ocs/index.php/WS/AAAIW16/paper/download/12566/12356>

6 September 2017

Nominet – Written evidence (AIC0131)

Nominet is a private internet company delivering public benefit, with a team of 150 people based in Oxford and London, and over 2,500 members who sell our domain names to businesses and the public. Nominet is responsible for one of the world's largest country code registries, running over 10 million domain names that end in .UK. Nominet also runs the Welsh Top Level Domains (TLDs) (.cymru and .wales), and provides registry services to a number of other branded and generic TLDs. Over 3 million businesses rely on Nominet's domain registry services.

In the 'machine learning' subcategory of artificial intelligence (AI) we've been applying sophisticated techniques to real-world applications in our everyday operations. The machine learning capabilities we have built into *turing*, a service that monitors the entirety of the UK domain name system (DNS), will allow it to learn from its past encounters and predictively respond to behaviour it sees on the web. For example, if it 'learns' through analysis of a data feed or human interaction that a certain type of behaviour is associated with a DDoS attack, it could a) decide to block all similar DNS requests meeting certain criteria and b) evaluate what those criteria should be by monitoring inputs from across the web.

Nominet has also applied machine learning to a recent domain categorisation project. After manually tagging a number of domains with their business type – by manually and painstakingly reviewing the website and making a decision about what type of business it represented – an algorithm we developed reviewed what made those domains receive those categories and was able to extrapolate and refine those judgements to the rest of the .UK namespace, using machine learning techniques. As a result we have more accurate insight into what .UK registrants are using their websites for than possibly any other registry in the world.

It is in this context that we are using our R&D function to help us capitalise on the learnings from running the DNS infrastructure into supporting new AI ideas and initiatives. For example, we are excited to be part of a co-operative of likeminded organisations who are exploring how we develop a robust framework to support the application of autonomous vehicles in a new project called DRIVEN. With £8.6 million in funding from the government, the DRIVEN project is set to be one of the first trials of a fleet of level 4 autonomous vehicles in the UK, testing cars that are considered to have 'high automation' and require a driver to be present, but without hands on the wheel or eyes on the road. Clearly, AI will play a significant role in the future deployment of connected and autonomous vehicles (CAVs). As experts in data analytics, Nominet will contribute to the project by creating secure platforms for real-time transactions and secure cloud infrastructure for data exchange. It is paramount that data collected from the car is protected with robust privacy and security protocols to prevent infiltration that could result in accidents.

We welcome the opportunity to respond to the Select Committee's public call for evidence on the economic, ethical and social implications of advances in artificial intelligence.

Ethics

AI is one of the most incredible advancements of our time and has now reached a point at which our society needs to prepare for application on a wide scale. This involves asking the big ethical questions and ensuring AI stays within moral frameworks that have yet to be designed.

This is most pertinent for AI that is equipped with machine learning; algorithms are employed to allow a computer to adapt over time in response to stimuli – or 'learn' from its interactions. With these forms of AI, the machine is striving to achieve a human-generated goal but can take whatever route it chooses to achieve it. This could result in potentially unpredictable or unintended consequences. A public debate needs to take place so that the appropriate frameworks are designed to ensure the route and choices made by AI are ethical.

Hard-wiring a complicated ethical code into a machine is a serious challenge for any software developer, especially as this decision-making process could make them liable for the consequences. The issue has been brought up often in discussions around autonomous vehicles. What will, and should, a car do in a situation when only one of two lives can be saved – pedestrian or driver? Who makes that decision and who is responsible for the consequences?

Transparency and trustworthiness will be essential components in any framework that ensures AI works with humans rather than replacing them. There needs to be a clear proof of the systems and workings of the AI to facilitate an investigation when mistakes are made. If we can't identify why AI did something, we can't make sure it doesn't repeat it.

Equally important is cooperation between the parties involved at every step of an AI machine's design, creation and application. Ethics needs to be considered at the point of creation, entwined in the workings rather than applied in retrospect. This will be complicated by the commercial nature of AI development and the swift advancements in technology, not to mention the challenge of 'codifying' a set of beliefs that all involved can agree on.

It is likely that a reliance on AI will change human behaviours and interactions, with unforeseen consequences. We also need to tackle security issues, bias, and potentially even the right of the robots with 'cognition'. Should they have the same rights as humans? Are they moral beings?

Cyber Security

When applied to network analytics, machine learning becomes a powerful weapon in the fight against cyber-attack and is a necessary component of any business' cyber security efforts. It offers improved flexibility and efficiency, with tools able to analyse the network in real time, live, without human oversight. Enterprise networks aren't neat and structured environments where a simple security policy is enough to deliver protection, but a dynamic place where patterns and threats emerge and evolve on an ongoing basis. The personalised insight, accuracy and speed of network analytics machine learning facilitates is vital as cyber criminals work harder to refine their methods of attack – indeed the criminals will likely use AI themselves in due course, with some experts suggesting they already do.

At Nominet, we have developed and use a DNS analytics tool called *turing*, the latest version of which has machine learning built-in to provide clear, actionable intelligence to drive strategies for protection. Businesses often lack the time and expertise for network analytics but recognise its importance to their cyber security efforts, so tools such as this can provide the information needed to mitigate threats, identify and solve any infrastructure issues, and even collate information to help identify and prosecute attackers.

In the years to come, machine learning will move from an exciting new development to a necessity for enterprise when it comes to cybersecurity, as the risks and costs of attacks grows. As more of our lives move online, we need to improve our defences, and recognise that human cyber security efforts will remain insufficient without the intervention of AI.

Impact on Society

One of the most immediate considerations when evaluating the future of any new technology is where it will take us as a society. One major challenge will be that of reskilling people, which will not be trivial. There will be those in low-skilled careers, those working in vocational trades, and the threats of technologies may have material implications on their career options. And many may not have the capacity or desire to adapt or change.

Without careful consideration, there is a risk that the digital divide will widen, between those who embrace AI and those ill-equipped to acquire the skills needed to pivot their careers in the precise moment that their role was replaced by a piece of AI-powered technology.

The Role of Government

It is possible that, further into the future, the machinery of government will need to fundamentally retool as the traditional revenue driven by taxation becomes harder to apply in a world of artificial machinery and intelligence. Consider the fact that a factory populated by a thousand workers pays a thousand workers' worth of income tax. But a factory equipped by 100 robots pays none, and may have the same output. Consideration will need to be given to future alternative revenue sources for government or we will face civil funding crises across the world.

Careful consideration should be given to the role of regulation and legislation, not necessarily of AI, but perhaps in relation to the obligations of organisations in responding to a new, AI reality. What commitments we need to make for reskilling and redeploying staff whose jobs were displaced by AI platforms, for example. Through all this, there will be a clear set of responsibilities the commercial sector will have to shoulder to drive a transition that allows all to benefit from the potential AI offers.

Industry

A further consideration is the implication of AI on business and entrepreneurship. A core premise of the internet for the last 20 years has been its ability to level the playing field. Anyone with a laptop and a connection could start up a website and a business in moments.

In a future defined and driven by AI, having the data, and having the sophisticated algorithms required to capitalise on it, will represent a huge set of digital assets a start-up might need access to, and therefore a barrier to entry for businesses looking to break into a new space. We must hope and encourage data sharing and code-open-sourcing to ensure that the advances we make in AI aren't limited to and held exclusively by the major established corporations, but rather that there is a shared knowledge base of algorithmic insight that can be made available to each successive generation of innovators and entrepreneurs.

6 September 2017

Norton Rose Fulbright LLP – Written evidence (AIC0079)

Background

1. The Select Committee on Artificial Intelligence (**Committee**) was appointed by the House of Lords on 29 June 2017 to consider the economic, ethical and social implications of the advance in artificial intelligence (**AI**). On 19 July 2017 it published a public call for written evidence (**Call for Evidence**), to be submitted to the Committee by 6 September 2017.
2. Norton Rose Fulbright is a global law firm. It provides the world’s pre-eminent corporations and financial institutions with a full business law service. It has more than 4,000 lawyers and other legal staff based in Europe, the United States, Canada, Latin America, Asia, Australia, Africa, the Middle East and Central Asia. The firm recently launched a website dealing with the legal and ethical aspects of AI: <http://aitech.law/>

This submission

3. As suggested by the Committee, this submission does not seek to answer all questions set out in the Call for Evidence. Rather, it addresses the following:

Question	Clarification provided by the Call for Evidence
3. How can the general public best be prepared for more widespread use of artificial intelligence?	<i>In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.</i>
7. How can the data-based monopolies of some large corporations, and the ‘winner-takes-all’ economies associated with them, be addressed? How can data be	

<p>managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?</p>	
<p>8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?</p>	<p><i>In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.</i></p>
<p>9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?</p>	

Our response

Question 3: How can the general public best be prepared for more widespread use of artificial intelligence?

4. Here we focus on some of the more narrow points suggested as being relevant to this question by the Committee, namely cyber security, privacy, and data ownership.
5. *Data ownership:* most products involving machine learning or AI rely heavily on proprietary datasets that are often not released. Keeping the data sets proprietary can provide implicit defensibility against competitors. An AI developer might seek to rely perhaps on the law relating to trade secrets/confidentiality and other intellectual property rights in relation to the data sets. The proprietary nature of its data sets may be particularly important to an AI developer in circumstances where open source software might otherwise lower barriers to entry by competitors. Where open data sets:
 - *are used to train AI:* the characteristics of the data sets may be more readily understood by the general public, with implications for an AI system reliant on such data sets accordingly more apparent and able to be scrutinised. Trust in the underlying data sets may help in establishing trust by the general public in an AI solution reliant on them. A number of open data sets are currently available (in both the public and private sectors). In the future it is possible that regulators may wish to look at steps to promote the development of open data sets; and

- *are not used*: proprietary data sets may contain inherent biases not obvious to the general public. The type of bias at issue may vary according to, say, the country, culture, or age demographic from which the data set was sourced.
6. *Privacy*: new data privacy laws, such as the EU General Data Protection Regulation (**GDPR**), are beginning to deal with AI explicitly. Under such privacy laws, key issues affecting the general public include whether:
 - all personal information used by an AI system has been collected with the data subject’s valid consent; and
 - such consent covers all purposes for which the AI uses the information.
 7. Typically, AI systems require large amounts of data to make intelligent decisions. Although some of this information alone might not be considered to be personal data, a large amount of it in combination with different data sources might make it possible to identify an individual and so breach data privacy laws. The general public do not have sufficient awareness of risk this currently.
 8. The use of AI raises data protection issues in relation to profiling. As users of AI systems frequently struggle to understand how such systems arrive at particular outputs or decisions, it is likely they will find it difficult to meet obligations under data privacy laws as regards:
 - notice;
 - other information provision requirements; and
 - providing a meaningful description of the relevant program logic explaining how an output or decision was reached (generic descriptions may not meet the criteria of new data privacy laws).
 9. It comes as no surprise, therefore, that a number of organisations working in the field of AI expect there to be an increasing need for rationale and explanation as to how the decision was reached by an algorithm. Assessment of how an algorithm reached a decision will in turn lead to scrutiny of how the algorithm was specified, designed, created, tested, verified, and calibrated before and after deployment. The availability of explanations as to how an algorithm reached a decision is likely to be an important factor in acceptance of the technology by the general public.
 10. Data privacy laws may give rise to other hurdles that will need to be addressed in the context of AI and personal data. For example, the GDPR provides that (subject to a few exceptions) individuals “shall have the right

not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her". As AI systems become more complex and draw on increasingly vast amounts of data, it may become very challenging for those operating AI systems to explain to a member of the general public how a decision was actually reached.

11. A business should therefore consider whether it would be useful or appropriate for a human to have the power to intervene before a decision by an AI system is finalised in order to check the decision, or to have the ability to override the decision afterwards.
12. New data privacy laws increasingly emphasise accountability and the idea of privacy by design and default. Designing AI to meet data privacy requirements from the outset is likely to be more cost-effective in the long run. This could include, for example, designing AI systems that can generate logs demonstrating what data was considered and what factors were taken into account.
13. *Cyber security*: where AI comes to be regarded as integral to critical national infrastructure or national security, government and security agencies may in time wish to have access to, say, data generated by AI, its algorithms, decision outputs, and explanations, and may use existing or new laws to obtain such access. In implementing such arrangements, government will need to be aware of the need to explain to the general public why such access is necessary.
14. As with any IT system, AI is susceptible to cyber intrusion and "hacking". Given the potential for both economic loss and physical damage (for example, AI-enabled robotics, or AI-controlled infrastructure) as a result of such incidents, businesses will need to ensure that all appropriate steps are implemented to: (1) guard against such risks; and (2) mitigate any breaches, in each case in accordance with applicable legal requirements and industry practices. Public confidence in AI technologies could be significantly eroded without such safeguards.

Question 7: How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

15. We answer this question from a legal perspective by reference to the competition law / antitrust considerations it raises.
16. See our submission under Question 3 in relation to data ownership.

16. Antitrust authorities, including the European Commission, are increasingly sensitive to the risk that AI, and algorithms more generally, can play a part in antitrust violations.
17. In recent years, the Commission and other European competition authorities have become increasingly concerned about the competitive implications of the growing use of algorithms. While the authorities recognise the pro-competitive uses of algorithms to help consumers find the lowest prices, they have already raised a number of concerns.
18. These concerns include the possibility that automated systems can help make price-fixing more effective, for instance by helping monitor deviations from price-fixing agreements or to implement price-fixing agreements in the first place, or facilitating other antitrust violations.
19. The growth of AI is likely to exacerbate these concerns. More speculatively, competitors' independent use of AI may lead to parallel behaviour without any coordination. Such behaviour is not currently prohibited, but can be expected to attract increasing antitrust scrutiny.

Question 8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

20. For AI systems to be accepted for use in a given market, as a matter of commercial reality the use of AI will need to be perceived by the participants in that market as meeting certain minimum ethical standards. Because legal and ethical responsibility are inextricably linked, that same commercial reality in effect imposes an imperative on businesses to address the corresponding legal issues (quite apart from a desire to limit risk).
21. *Why are ethics a key issue for the AI industry:* AI systems are typically both the outcome of, and result in, a movement of ethical decision-making to an earlier stage in a system's life-cycle. This can have profound implications for where ethical and legal responsibility can lie in a given AI supply chain.
22. *Can human values be embedded in AI?:* the idea that AI systems should be designed at inception to embed human values in order to avoid breaching human rights and to avoid creating bias, commonly known as "ethics-bounded optimisation", is increasingly accepted within the AI industry as a way of mitigating risk. However, AI will not change the fact that those who breach legal obligations in relation to human rights will still be responsible for such breaches (although it may make determining who is responsible more complex).

23. Addressing risk by attempting to embed human values in AI systems may be extremely difficult for a range of reasons, not least because the definition of what is a societal norm may differ over time, between markets, and between geographies. Social norms may also differ between communities and sub groups of the population.
24. *What steps should be taken to minimise the risk of bias?*: designers, developers, and manufacturers of AI will wish to avoid creating unacceptable bias from the data sets or algorithms used. To mitigate the risk of bias, they will need to understand the different potential sources of bias, and the particular AI system will need to integrate identified values and enable third party evaluation of those embedded human values to detect any bias.
25. *How can transparency be achieved?*: AI and AI-enabled products and services will need to incorporate a degree of ethical transparency in order to engender trust (otherwise market uptake may be impeded). This will be particularly important when AI autonomous decision-making has a direct impact on the lives of market participants. How can such ethical transparency be achieved? There are two separate elements. AI should:
- be open, understandable and consistent to minimise bias in decision-making; and
 - deliver transparency as to the decision or action.
26. *What steps are required to achieve accountability?*: legal systems will need to consider how to allocate legal responsibility for loss or damage caused by AI systems. As they proliferate and are allowed to control more sensitive functions, unintended actions are likely to become increasingly dangerous. There should accordingly be program-level accountability to explain why an AI system reached a decision to address questions of legal responsibility.
27. *Inserting humans in the loop*: the complexity of AI systems in combination with emerging phenomena they encounter mean that constant monitoring of AI systems, and keeping humans “in the loop”, may be required. However, while keeping humans “in the loop” may help to achieve accountability, it may also limit the intended benefits of autonomous decision-making. A balance will need to be struck.
28. Legislative initiatives are being considered in a number of jurisdictions to address questions of accountability. These include a registration process for AI, identity tagging, criteria for allocating responsibility, and an insurance framework.

Question 9: In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

29. So-called AI black boxing and lack of AI transparency are likely to be highly problematic from a legal and ethical perspective where AI systems:

- are used in safety-critical situations;
- could: (1) produce legal effects on a human; or (2) affect a human's rights, freedoms or legitimate interests; or
- could otherwise significantly affect a human to its detriment.

30. Designers, developers, and manufacturers may wish to protect their intellectual property rights in an AI system as a trade secret, which can lead to a deliberate lack of transparency in such systems.

31. Where lack of transparency is an issue, a number of alternatives could be considered, including:

- so-called "interactive machine learning". This puts interactions with humans as a central part of developing machine learning systems. It includes building in functionality that enables: (1) an AI system to "explain" its decision-making to a human; and (2) the human to give feedback on the system's performance and decision outcomes. However, while keeping humans "in the loop" may help to achieve accountability, it may also limit the intended benefits of autonomous decision-making. Not having to involve humans may have been the reason the AI was implemented in the first place; and
- implementing a process (as part of the design and build of a system) to automate a log or report for a human user of operations and decisions to enable audits and increase transparency. Any assumptions relied on should also be included in the log. The logic and rules of AI systems should also be available as needed.

We hope this this submission is useful for the Committee. We are happy to discuss the above submission with it in more detail.

Mike Rebeiro

Partner and Global Head, Technology and Innovation

Norton Rose Fulbright LLP

5 September 2017

Professor Thomas Nowotny, Dr Andrew Philippides, Dr Paul Graham and Professor James Marshall – Written evidence (AIC0088)

Professor Thomas Nowotny, Dr Andrew Philippides, Dr Paul Graham and Professor James Marshall – Written evidence (AIC0088)

[Submission to be found under Professor James Marshall](#)

NVIDIA – Written evidence (AIC0212)

Table of Contents

	<u>Introduction</u>
1	<u>Current State of Artificial Intelligence</u>
	<u>Past</u>
	<u>Present</u>
	<u>Contributing Factors for the Modern AI Age</u>
	<u>Future</u>
	<u>Factors affecting future development</u>
2	<u>Level of Excitement Surrounding AI</u>
3	<u>Preparing the General Public</u>
4	<u>Societal Gains and Losses</u>
5	<u>Improving the Public’s Understanding</u>
6	<u>Key Sectors that will Benefit from AI</u>
7	<u>Data-Based Monopolies</u>
8	<u>Ethical Implications</u>
9	<u>Transparency of AI Systems</u>
10	<u>The Government’s Role</u>
11	<u>Learning from others</u>
12	<u>Appendix A – References</u>
13	<u>Appendix B – Tables and Diagrams</u>

INTRODUCTION

NVIDIA would like to thank the members of the Select Committee for AI and the Committee Secretarial team for the opportunity to provide evidence for this important enquiry.

NVIDIA is an acknowledged thought-leader in AI, as well as the creator of many of the hardware and software components that combine to provide the modern AI Platform. As such, we are frequently called upon by committees around the world to provide insight and direction in AI programs.

This response has been collated by members of our Research Computing team, consisting of Supercomputing specialists, AI subject-matter experts and NVIDIA representatives covering specific industry verticals such as AI in Healthcare, Autonomous Vehicles and Robotics. We use this experience to separate reality from fiction, provide a ground-truth on what modern AI can do and, in the space permitted, an insight into where the technology is taking us.

1. CURRENT STATE OF ARTIFICIAL INTELLIGENCE

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Past

Born out of the Dartmouth Conference, 1956, artificial intelligence is a very broad field of research with hundreds of approaches and applications. This decade has seen considerable progress especially in one branch of artificial intelligence, which is deep representation learning, with deep neural networks being one of the most successful algorithm families from this branch [1].

It is important to note that among the many types of artificial intelligence, the key technology that has created the modern AI boom is the neural network. Legacy technologies will remain useful in many domains, but any examination of the effects and opportunity rising from the AI revolution should focus on neural networks.

Neural networks are not a new algorithm; the theory was coined in the 1940s [3; 4] and first implementations appeared in the '50s [2] (full historical perspective in [1]). However, the technology was immature and the development of classical machine learning models and especially kernel machines [5; 6; 7] significantly surpassed neural networks in performance. This resulted in the decline in popularity of neural networks (“AI winter”) which lasted until 2006/2007 [8; 9; 10].

Present

Since 2007, neural network performance changed from the position where they were considered as a toy concept incapable of solving real-world problems, to a stage where they can surpass expert human performance for a large number of

complex tasks such as: image recognition [11]; speech recognition [12]; relational reasoning [13]; some simple games like chess, Go or Atari games [14; 15]; complex games like DOTA 2 [16] (work towards Starcraft 2 is in progress [17]); a range of medical tasks like cancer detection [18]; face recognition [19]; agricultural produce recognition [20; 21] and many more. Combined with the fact that computers can execute algorithms at a scale not feasible for humans (think about being able to look at all CCTV cameras in a city continuously while detecting accidents from each in real-time) this creates an almost infinite amount of new technological opportunities (see question 2).

Moreover, research is suggesting that we have not reached the limits of neural network performance and that performance will continue to grow logarithmically with the order of magnitude of the datasets [22].

Contributing Factors for the Modern AI Age

The recent revolution in neural network performance was not caused by a significant algorithmic breakthrough. Instead, there were three other factors that allowed the technology to progress [1]:

1. **Data:** The size of available datasets increased considerably (because of decreasing cost of storage and internet revolution) which allowed the use of larger models [Diagram 1 in Appendix B]. The age of AI would have not been possible without the age of Big Data.
2. **Model size** and complexity increased considerably, which allowed the performance to increase but required n^2 the compute of smaller models [Diagram 2 in Appendix B].
3. **Computing power** has increased proportionally to the model size and data size increase; for example, the latest NVIDIA DGX-1V computer (which is 3 rack units or 5¼ inches high) will deliver 960 TFLOPs (10^{12} floating point operations per second) which is equivalent to the IBM Roadrunner supercomputer [40] (occupying a sizable building) which was the most powerful system in the world in 2008.

The key computing component of the modern AI revolution is the NVIDIA graphics processing unit (GPU). The ability to program a commonly available GPU for mathematical problems other than traditional graphics has enabled researchers everywhere to deploy previously unthinkable computing performance to their desktop.

Today, the world's most powerful AI Supercomputers are based at centres combining expert capabilities in GPU computing and neural networks [23]. The progress in GPU computing has not only delivered a dramatic increase in computational density but also at a fraction of energy cost of traditional hardware (highest performance per watt of energy).

Future

Even though we have seen the unprecedented success of neural networks, our understanding of this technology is still in its infancy; therefore, it is difficult to predict what shape it will take.

► One thing is certain though, to quote Andrew Ng (Professor at Stanford University and former Chief Scientist at Baidu):

1. *"...today's supervised learning software (the software which requires data that is annotated by a human) has an Achilles' Heel: It requires a huge amount of data." [24]. That is expensive to curate and validate. Therefore, the research community widely agrees that "the next Frontier in AI" is unsupervised / semi-supervised / predictive learning [25].*

The future of AI will operate with less human labelled data, or it will not require it at all. These developments will have a fundamental impact on the technology adaptation as the dataset will cease to be a barrier to entry, democratising the technology adaptation. It will also allow for the use of datasets orders of magnitude bigger than those used today, which will lead to significantly increased accuracy (at the cost of computation). It will also lead to adaptation of AI to areas where datasets are very difficult or costly to collect (e.g. some areas in medicine).

We are already seeing considerable progress in this area:

1. Generative adversarial networks [26] being called the most important discovery of neural network research this decade [27] with a wide range of other generative approaches being investigated [44].
2. Semi-supervised learning approaches are successfully improving AI performance [28].
3. Neural networks' resilience to labelling noise proven to be significant (even 20% noise not impacting performance in some applications) [22].

Another key area of research looks at the problems that require a level of planning or interaction with other agents (whether people or AI). Those problems are being addressed by a family of reinforcement learning algorithms (which also at their backbone are built from neural networks). This is the technology underpinning autonomous vehicles, advanced robotics in manufacturing or surgery, a military force multipliers, or even energy management.

Reinforcement learning is a technology in its infancy, through with AI algorithms already surpassing human ability for simple games like chess, Go or Atari games (simple since the AI has all the information and the set of rules is very constrained) but also for more complex games like Dota 2 [16] or Starcraft 2 [17]. Significant effort is put into the research of relational reasoning, e.g. the relationships between objects and events [34], learning by example [35] and replicating processes like human imagination for training neural networks, e.g. to learn from your imagination without experiencing the actual event [33].

This technology will allow us to address an even wider set of problems with solutions such as NVIDIA ISAAC [42], a photorealistic virtual world obeying the laws of physics that can be used for training robots by simulating real life at the speed of silicon.

The progress in the neural network research is unprecedented with much significant research being published every month. Due to recommendations on

submission length it is not possible to cover many of them in detail. Other areas that deserve attention soon include:

- Federated learning [46] – that is, the ability to train neural networks in a distributed way on private data without the need for the data to leave the owner. This also includes work on blockchain implementation of the above [45, 47].
- Meta-learning – that is, the ability for neural networks to design themselves [48].
- Common sense [50] and learning to reason [49].

Factors affecting future development

The technology has reached a level of maturity and adaptation where very few things will hinder future progress (refer to question 11). The more relevant question is whether the UK will play a leading role in the development of AI, and if it doesn't, what impact will this have on the UK's economic, social and military influence.

The technology specific point of view on the factors that will impact UK leadership in AI are discussed under question 3.

2. LEVEL OF EXCITEMENT SURROUNDING AI

Is the current level of excitement which surrounds artificial intelligence warranted?

In the last decade, we have moved from the position where AI algorithms like neural networks didn't work for meaningful tasks to a position today where: "If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future." [24]. Moreover, AI algorithms can achieve this at a speed and scale only possible for computers.

What makes this technological opportunity so exciting for the research and business community is that an overwhelming proportion of our daily lives is spent on simple, repeatable tasks that we no longer need to apply direct concentration to. An experienced driver does not have to think about the process of driving. An experienced medical consultant can look at a medical image and immediately infer the critical facts based upon previous learnings. A seasoned translator can convert a sentence from French to English entirely with unconscious thought.

Other examples include a manufacturing quality assessment, infrastructure inspection, finding goods in warehouses, counting blood cells in a microscope sample, sorting grains of rice from stones and sticks, inspecting CCTV cameras for accidents and acts of vandalism, tracking crop harvest from satellite imagery, inspecting large construction sites for health and safety issues or lapses in quality, tracking freighters in busy shipping lanes. This list is commensurate with the level of excitement AI warrants.

That is why the key figures in the AI research community are starting to compare AI to electricity:

"Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI will transform in the next several years, ..." [32]. — Andrew Ng.

3. PREPARING THE GENERAL PUBLIC

How can the general public best be prepared for more widespread use of artificial intelligence?

AI will extend into all areas of life, globally. With all technological revolutions, there is an opportunity to exploit the technology for maximum societal gains. Other commentators will provide insight to their specific area of expertise.

- ▶ From NVIDIA's perspective, the single greatest way to ensure that the United Kingdom gains from the modern AI revolution is through education.
- ▶ Special emphasis should be placed on:
 1. Revisiting the primary school curriculum (it is possible to deliver computer programming from the age of 8 years [36], and shortly after introduce artificial intelligence combined with robotics [37]).
 2. Revisiting the secondary school curriculum (building on the above and accelerating introduction of linear algebra and elements of calculus).
 3. Revisiting the higher education curriculum to significantly extend the AI curriculum on computing courses, but also making sure there is an introduction to AI for most programs with program specific applications, e.g. AI in medicine, AI in finance, AI in social science.
 4. Establishing national computational infrastructure and ensuring it is widely available for education at all levels.
 5. Promoting AI research at all levels in multiple scientific domains as a general purpose scientific enabling technology.

4. SOCIETAL GAINS AND LOSSES

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Due to recommendations on submission length, NVIDIA will focus our response on a handful of topics. We remain open to providing evidence on this subject later if required.

5. IMPROVING THE PUBLIC'S UNDERSTANDING

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

We refer to our response to question 3. Public understanding is best served through early adoption of the technology across the school system. Otherwise, due to recommendations on submission length, NVIDIA will focus our response

on a handful of topics. We remain open to providing evidence on this subject later if required.

6. KEY SECTORS THAT WILL BENEFIT FROM AI

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Due to recommendations on submission length, NVIDIA will focus our response on a handful of topics. We remain open to providing evidence on this subject later if required.

7. DATA-BASED MONOPOLIES

How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Due to recommendations on submission length, NVIDIA will focus our response on a handful of topics. We remain open to providing evidence on this subject later if required.

8. ETHICAL IMPLICATIONS

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

Due to recommendations on submission length, NVIDIA will focus our response on a handful of topics. We remain open to providing evidence on this subject later if required.

9. TRANSPARENCY OF AI SYSTEMS

In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

The assertion that neural networks lack transparency is false. The nature of neural networks is different from that of hand-written computer programs; therefore, new methods for validation must be, and indeed are being, developed [38, 39, 51, 52].

Deep neural networks are just complex mathematical functions, approximating the relationship between datasets (e.g. finding a mapping between English and French). Neural networks are complex because they have millions or billions of parameters and because of that vastness are not practical for a human to read. From this perspective, they are no different from manually written source code. Reading and understanding two lines or even a hundred lines of code is practical (similarly to a tiny neural network) but when the complexity increases to thousands or hundreds of thousands of lines of code it impractical or impossible

to inspect it manually. WindowsXP contained approximately 45 million lines of code, impossible for any single human to comprehend.

In practice, other code validation methods are used (for example unit or coverage tests) to ensure source code is functioning correctly and is secure (security breaches are just lapses in that test methodology). The fact that the code is written in a programming language just provides an illusion of transparency as it is not possible for humans to read and the only validation can be done through thorough tests. From this perspective, neural networks offer significant advantage over the hand-written code. Not only do the standard test procedures still apply (as it is common practice to test them against a test dataset), a neural network's code complexity is negligible (to the extent where the code can be easily read and understood by a single human) and more importantly they can be more formally tested and their performance mathematically proven (since they are just complex functions).

10. THE GOVERNMENT'S ROLE

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

The Government must prepare the general public as described in our response in Section 3.

The Government must maintain the entrepreneurial environment required to stimulate both business and universities not only to push the boundaries of AI research, but also to exploit this research in the increasingly competitive market that the United Kingdom finds itself in. The United Kingdom has the opportunity to provide a legislative umbrella making the country a desirable place to develop, test and deploy the latest AI algorithms.

Investment in modern methods of AI has been visibly below that of the UK's competing nations [29].

11. LEARNING FROM OTHERS

What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

A key difference in the approaches of other nations toward fully exploiting developments in modern AI is that of funding. We note the Chinese government has announced a plan to grow the Chinese AI economy to \$150 billion by 2030 [29], with a goal of catching up with U.S. AI development by 2020 (not unrealistic since companies like Baidu are already world leaders in AI). This involves a very wide range of initiatives including \$5 billion Tianjin government investment into AI [29]. We can see similar situations across the globe.

The South Korean government has announced it would spend 1 trillion won (USD 840 million) by 2020 to boost the artificial intelligence industry. Taiwan's Ministry of Science and Technology undertakes five strategic tasks with a 5-year budget of NT\$16 billion (USD 527 million) to boost the country's development of AI-related industries and applications. This includes building up high performance computing capabilities, setting up R&D centers for AI innovation and smart robotics innovation bases, additional budget for semiconductor makers as well as student challenges to stimulate AI technology innovation [54]

Canada is investing C\$ 125M in a national AI strategy to be spent at a number of leading research institutions, with Vector being one. [55]

It is believed that Baidu and Google alone have invested between 20 to 30 billion dollars in AI research [30]. Microsoft listed AI as one of its top priorities [31] and companies like NVIDIA, Facebook, Amazon, Intel, IBM, Netflix are investing huge proportion of their revenue in infrastructure, staff and acquisitions.

12. APPENDIX A – REFERENCES

1. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
2. Rosenblatt, Frank. "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological review* 65.6 (1958)
3. McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." *The bulletin of mathematical biophysics* 5.4 (1943)
4. Hebb, Donald. "0. (1949)." *The organization of behavior* (1957).
5. Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In COLT '92: Proceedings of the fifth annual workshop on Computational learning theory, pages 144–152, New York, NY, USA. ACM.
6. Cortes, C. and Vapnik, V. (1995). Support vector networks. *Machine Learning*.
7. Schölkopf, B., Burges, C. J. C., and Smola, A. J. (1999). *Advances in Kernel Methods — Support Vector Learning*. MIT Press, Cambridge, MA.
8. Hinton, G. E., Osindero, S., and Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18, 1527–1554.
9. Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2007). Greedy layer-wise training of deep networks. In NIPS'2006
10. Ranzato, M., Poultney, C., Chopra, S., and LeCun, Y. (2007a). Efficient learning of sparse representations with an energy-based model. In NIPS'2006
11. He, Kaiming, et al. "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification." *Proceedings of the IEEE international conference on computer vision*. 2015.
12. Amodei, Dario, et al. "Deep speech 2: End-to-end speech recognition in English and mandarin." *International Conference on Machine Learning*. 2016.
13. Santoro, Adam, et al. "A simple neural network module for relational reasoning." *arXiv preprint arXiv:1706.01427* (2017).
14. Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." *Nature* 529.7587 (2016): 484-489.
15. Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." *arXiv preprint arXiv:1312.5602* (2013).
16. "More on Dota 2", OpenAI blog (16 August 2017), Retrieved from: <https://blog.openai.com/more-on-dota-2/>
17. Vinyals, Oriol, et al. "StarCraft II: A New Challenge for Reinforcement Learning." *arXiv preprint arXiv:1708.04782* (2017).
18. Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542.7639 (2017): 115-118.
19. Nakada, Masaki, Han Wang, and Demetri Terzopoulos. "AcFR: Active Face Recognition Using Convolutional Neural Networks." *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017 IEEE Conference on. IEEE, 2017.
20. Kussul, Nataliia, et al. "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data." *IEEE Geoscience and Remote Sensing Letters* 14.5 (2017): 778-782.

21. Chostner, Ben. "See & Spray: The Next Generation of Weed Control." *Resource Magazine* 24.4 (2017): 4-5.
22. Sun, Chen, et al. "Revisiting unreasonable effectiveness of data in deep learning era." *arXiv preprint arXiv:1707.02968* (2017).
23. "Green Top 500 list", August 2017, Retrieved from: <https://www.top500.org/green500/list/2017/06/>
24. "What Artificial Intelligence can and cannot do right now", Andrew Ng, November 2016, Retrieved from: <https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now>
25. "(CMU RI seminar) Yann LeCun: Unsupervised Learning: The Next Frontier in AI", Yann LeCun, November 2016, Retrieved from: <https://www.youtube.com/watch?v=XTbLOjVF-y4>
26. Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.
27. "What are some recent and potentially upcoming breakthroughs in deep learning?", Yann LeCun, July 2016, Retrieved from: <https://www.quora.com/session/Yann-LeCun/1>
28. Laine, Samuli, and Timo Aila. "Temporal Ensembling for Semi-Supervised Learning." *arXiv preprint arXiv:1610.02242* (2016).
29. "Beijing Wants A.I. to Be Made in China by 2030", July 2017, Retrieved from: <https://www.nytimes.com/2017/07/20/business/china-artificial-intelligence.html?mcubz=0>
30. "McKinsey's State Of Machine Learning And AI, 2017", July 2017, Retrieved from: <https://www.forbes.com/sites/louiscolombus/2017/07/09/mckinseys-state-of-machine-learning-and-ai-2017/#397002b775b6>
31. "Microsoft Bids Goodbye to 'Mobile First' Mantra in Favor of AI", August 2017, Retrieved from: <http://fortune.com/2017/08/03/microsoft-cloud-ai-mobile>
32. "Andrew Ng: Why AI Is the New Electricity", Andrew Ng, March 2017, Retrieved from: <https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity>
33. Weber, Théophile, et al. "Imagination-Augmented Agents for Deep Reinforcement Learning." *arXiv preprint arXiv:1707.06203* (2017).
34. Santoro, Adam, et al. "A simple neural network module for relational reasoning." *arXiv preprint arXiv:1706.01427* (2017).
35. Christiano, Paul, et al. "Deep reinforcement learning from human preferences." *arXiv preprint arXiv:1706.03741* (2017).
36. Carol Vorderman , "Help Your Kids with Computer Coding", DK Publishing (2014)
37. Laurens Valk, "LEGO MINDSTORMS EV3 Discovery Book: A Beginner's Guide to Building and Programming Robots", No Starch Press (2014)
38. Karpathy, Andrej, Justin Johnson, and Li Fei-Fei. "Visualizing and understanding recurrent networks." *arXiv preprint arXiv:1506.02078* (2015).
39. Huang, Xiaowei, et al. "Safety verification of deep neural networks." *International Conference on Computer Aided Verification*. Springer, Cham, 2017.

40. Military Supercomputer Sets Record, June 2008, Retrieved from: <http://www.nytimes.com/2008/06/09/technology/09petaflops.html>
41. Gehring, Jonas, et al. "Convolutional Sequence to Sequence Learning." *arXiv preprint arXiv:1705.03122* (2017).
42. "Virtual Simulator for robots", May 2017, Retrieved from: <https://www.nvidia.com/en-us/deep-learning-ai/industries/robotics/>
43. "AI Influencer Andrew Ng Plans The Next Stage In His Extraordinary Career", June 2017, Retrieved from: <https://www.forbes.com/sites/peterhigh/2017/06/05/ai-influencer-andrew-ng-plans-the-next-stage-in-his-extraordinary-career/#59bdc6d3a2ce>
44. Chen, Qifeng, and Vladlen Koltun. "Photographic image synthesis with cascaded refinement networks." *arXiv preprint arXiv:1707.09405* (2017).
45. "OpenMinded, encrypted, decentralised artificial intelligence", September 2017, Retrieved from: <http://openminded.org/>
46. Konečný, Jakub, et al. "Federated learning: Strategies for improving communication efficiency." *arXiv preprint arXiv:1610.05492* (2016).
47. Zyskind, Guy, and Oz Nathan. "Decentralizing privacy: Using blockchain to protect personal data." *Security and Privacy Workshops (SPW), 2015 IEEE*. IEEE, 2015.
48. Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks." *arXiv preprint arXiv:1703.03400* (2017).
49. Andreas, Jacob, et al. "Neural module networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.
50. Weston, Jason, Sumit Chopra, and Antoine Bordes. "Memory networks." *arXiv preprint arXiv:1410.3916* (2014).
51. Huang, Xiaowei, et al. "Safety verification of deep neural networks." *International Conference on Computer Aided Verification*. Springer, Cham, 2017.
52. "Federated Logic Conferences", September 2017, Retrieved from: <http://www.floc2018.org/conferences/>
53. "Diffblue, automated test generation", September 2017, Retrieved from: <http://www.diffblue.com/>
54. <http://english.yonhapnews.co.kr/news/2016/03/17/0200000000AEN20160317003751320.html>
55. <https://www.cifar.ca/assets/pan-canadian-artificial-intelligence-strategy-overview/>

13. APPENDIX B – TABLES AND DIAGRAMS

Diagram 1: Extract from [1] discussing the recent growth of available datasets

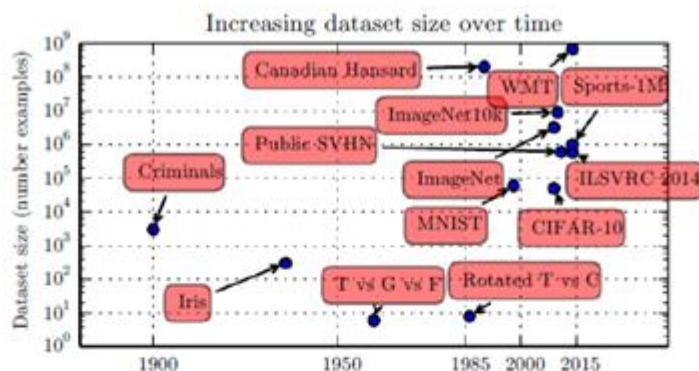


Figure 1.8: Dataset sizes have increased greatly over time. In the early 1900s, statisticians studied datasets using hundreds or thousands of manually compiled measurements (Garson, 1900; Gosset, 1908; Anderson, 1935; Fisher, 1936). In the 1950s through 1980s, the pioneers of biologically inspired machine learning often worked with small, synthetic datasets, such as low-resolution bitmaps of letters, that were designed to incur low computational cost and demonstrate that neural networks were able to learn specific kinds of functions (Widrow and Hoff, 1960; Rumelhart *et al.*, 1986b). In the 1980s and 1990s, machine learning became more statistical in nature and began to leverage larger datasets containing tens of thousands of examples such as the MNIST dataset (shown in Fig. 1.9) of scans of handwritten numbers (LeCun *et al.*, 1998b). In the first decade of the 2000s, more sophisticated datasets of this same size, such as the CIFAR-10 dataset (Krizhevsky and Hinton, 2009) continued to be produced. Toward the end of that decade and throughout the first half of the 2010s, significantly larger datasets, containing hundreds of thousands to tens of millions of examples, completely changed what was possible with deep learning. These datasets included the public Street View House Numbers dataset (Netzer *et al.*, 2011), various versions of the ImageNet dataset (Deng *et al.*, 2009, 2010a; Russakovsky *et al.*, 2014a), and the Sports-1M dataset (Karpathy *et al.*, 2014). At the top of the graph, we see that datasets of translated sentences, such as IBM’s dataset constructed from the Canadian Hansard (Brown *et al.*, 1990) and the WMT 2014 English to French dataset (Schwenk, 2014) are typically far ahead of other dataset sizes.

Diagram 2: Extract from [1] discussing the increase in size of modern artificial neural networks and comparison to biological organisms.

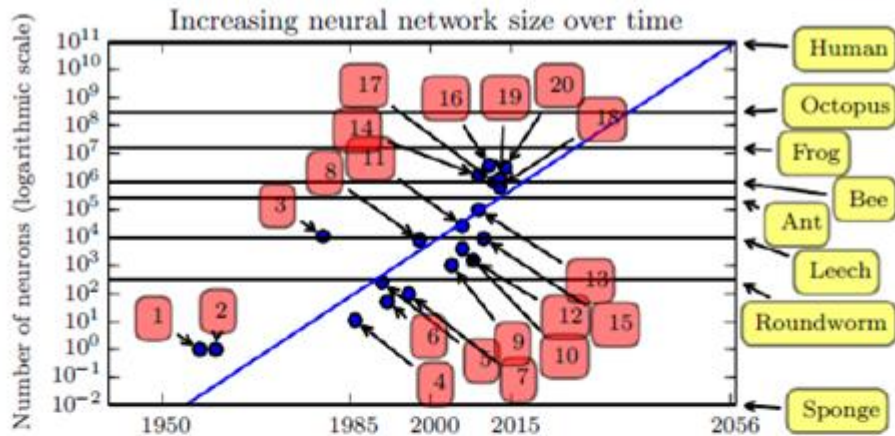


Figure 1.11: Since the introduction of hidden units, artificial neural networks have doubled in size roughly every 2.4 years. Biological neural network sizes from [Wikipedia \(2015\)](#).

1. Perceptron ([Rosenblatt, 1958, 1962](#))
2. Adaptive linear element ([Widrow and Hoff, 1960](#))
3. Neocognitron ([Fukushima, 1980](#))
4. Early back-propagation network ([Rumelhart et al., 1986b](#))
5. Recurrent neural network for speech recognition ([Robinson and Fallside, 1991](#))
6. Multilayer perceptron for speech recognition ([Bengio et al., 1991](#))
7. Mean field sigmoid belief network ([Saul et al., 1996](#))
8. LeNet-5 ([LeCun et al., 1998b](#))
9. Echo state network ([Jaeger and Haas, 2004](#))
10. Deep belief network ([Hinton et al., 2006](#))
11. GPU-accelerated convolutional network ([Chellapilla et al., 2006](#))
12. Deep Boltzmann machine ([Sriakbutdinov and Hinton, 2009a](#))
13. GPU-accelerated deep belief network ([Rains et al., 2009](#))
14. Unsupervised convolutional network ([Jarrett et al., 2009](#))
15. GPU-accelerated multilayer perceptron ([Ciresan et al., 2010](#))
16. OMP-1 network ([Coates and Ng, 2011](#))
17. Distributed autoencoder ([Le et al., 2012](#))
18. Multi-GPU convolutional network ([Krizhevsky et al., 2012](#))
19. COTS HPC unsupervised convolutional network ([Coates et al., 2013](#))
20. GoogLeNet ([Szegedy et al., 2014a](#))

12 September 2017

Ocado Group plc – Written evidence (AIC0050)

OCADO Response to Select Committee on Artificial Intelligence Call for Evidence

As Ocado expands, it is using its own technology, including artificial intelligence (AI) to ensure that our customers are served as effectively and efficiently as possible. Using the insights we have acquired as we continue to evolve, Ocado has provided a response to the House of Lords as it looks at the economic, ethical and social advances in AI.

THE PACE OF TECHNOLOGICAL CHANGE

1. What is the current state of artificial intelligence and what factors have contributed to this?

How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- We are about to witness a massive explosion of AI applications fuelled by:
 - Recent advances in AI component technologies such as deep learning, natural language processing, image recognition, convolutional neural nets etc
 - Reduction in the cost, power usage and physical footprint, and increase in the performance of, processors (and associated electronic components) to power embedded applications of AI
 - Advances in cloud computing:
 - Building AI into your applications used to require hiring specialist engineers with PhDs and specialist applications of AI still do. However for other applications, the advent of cloud-based services has completely disrupted accessibility to AI and made it a commodity, and in so doing has redefined the baseline for smartness
 - Now for a few cents you can call a cloud API, pass in some data and get back a smart prediction or insight. If applications are not taking advantage of this new smartness, then they will probably not be best of breed
 - As well as providing smart predictions or insights for a lower cost, cloud has also enabled companies to store and process at lower cost the data required to train their AI solutions
 - The alignment of a number of technological planets:
 - Where AI, Big Data, Robotics, Internet of Things (IoT) and Cloud Computing collide, we will see an exponential increase of smart mobile systems that are able to communicate with one another and interact with the world around them via IoT

- This is not a hypothetical prediction - this combination of technologies already powers our business at Ocado. Further details of this can be found at the end of the document
- It seems inevitable that GDPR is going to be a significant challenge to the evolution of AI because of how it will limit how data can be used to create smarter systems and insights:
 - Clearly privacy and the appropriate use of customer data are important issues
 - On the one hand our customers expect our systems to get to know them, to personalise the services we offer them, to predict their future needs, to reduce friction by enabling them to shop faster and so on
 - On the other hand customers can be offended by assumptions you make about them based on their previous actions, even if those assumptions are correct
- Another key hindrance is the digital skills gap - see 3) below
- Security is an incredibly important challenge facing both AI and IoT

2. Is the current level of excitement which surrounds artificial intelligence warranted?

- At Ocado we believe that this excitement is warranted. The impact of AI as a transformative force is definitely under-estimated
- What makes AI different from other disruptive technologies is that the disruption is recursive (existing AIs can be used to help train and optimise the next generation of AIs) and the fact that AI can help us discover things we did not know (knowledge mining)
- AI is also in a disruptive league of its own because it is a vital ingredient when you want to do anything really exciting with other disruptive technologies such as IoT, Big Data, Cloud and Robotics. When it comes to disruptive technologies, you could say, "AI is the one to rule them all"

IMPACT ON SOCIETY

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

- The UK is not yet preparing adequately the next generation for the smart digital world they will inhabit. As such, substantially more needs to be done to achieve a paradigm shift in the quality and quantity of STEM students, graduates and professionals over the long-term

- We believe the current school curriculum is not fit for purpose in relation to STEM. Indeed, there is a need for much more bold thinking by government about the complete educational pipeline to feed our digital talent pool, of which AI is obviously a key element
- Looking wider, we also need to prepare the next generation for living in a world populated by smart machines, digital assistants and robots:
 - The current school curriculum needs to be updated in line with not only today's demands but the demands of tomorrow too
 - Many of the skills and techniques we are currently teaching our children will be as devalued in the years to come just as the encyclopaedia has been by the internet. Instead we need to focus on teaching the meta skills such as collaboration, creative thinking, intersectional thinking, entrepreneurship etc
 - We also need to help our children understand the amazing possibilities and current limitations of AI, the important ethical and philosophical questions around the use of AI and so on
 - We need to change the perception of technology by girls at a young age in order to get more women into these technology industries, and help remedy the significant shortage of women with STEM qualifications
- Coding and data literacy should be mandated just like English and Maths are. They are essential transformative skills not just for those who may go on to become computer scientists but for everyone. They are also stepping stones towards literacy in AI. And obviously, mandating these subjects will significantly help to reduce the gender gap in STEM and the subsequent earnings disparity within the labour market
- We believe a new approach is required around the teaching of technology within primary and secondary schools. For example, the good intentions behind launching a primary computing curriculum are let down when the UK lacks the qualified teachers to deliver the programme. At a secondary level, although the introduction of a computer science GCSE is an important step, without schools being mandated to offer it to pupils, its impact will be dramatically weakened. Given the importance of early learning in relation to the subject matter, rhetoric must be matched by policy. Furthermore, we need to be exploring using AI to help teach those topics such as computer science where we lack sufficient skilled teachers
- As a company we decided to tackle the problem at source and, as a result, we developed the Code for Life initiative (<https://www.codeforlife.education/>), developed entirely in-house by our Ocado Technology team and rolled out as part of our Corporate Responsibility agenda
- Code for Life is a free teaching tool, and a fun game, designed to get young children coding. It is currently in use at over 1,000 schools in the UK, and has over 68,000 registered pupils learning to code
- Teaching children to code is just the first step towards true digital literacy. We also need to teach them to be data literate. To understand how to organise and manipulate data, to gain insights from them, to visualise them, to build

models from them and so on. We need to be weaving digital literacy throughout the curriculum

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

- The best way to address these potential disparities is through education and achieving widespread digital literacy, as discussed in 3) above

PUBLIC PERCEPTION

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

- Whenever the media touch on the subject of AI and robotics, it almost always focuses on negative outcomes such as Big Brother watching, "robots taking jobs" or a vision for AI Armageddon
- To restore the balance, foster greater adoption of these technologies and the skill sets that underpin them, there is a need to generate more positive stories
- Such stories include:
 - How we might use these technologies to augment human beings, to achieve things we currently cannot do, don't want to do or where we add little value. For example, helping doctors make better medical diagnoses and improve the outcomes of surgical procedures
 - How we might use AI to make smarter use of our scarce resources whether that be time, energy, water, land, transport network capacity etc
 - How AI can help improve our physical and cyber security, including how smart agents can help us to safely navigate a world awash with sensors and data on the back of IoT
 - How we might use AI to offset or hopefully reverse the impact of climate change, pollution and poverty
 - How we might use AI to generate insights, manage complexities and make discoveries that are beyond our human minds. For example, analysing medical scans or data from wearables to identify the signatures of medical conditions in their early stages
 - How we might use AI to tackle major societal challenges such as remote healthcare and providing care (and even companionship) to our growing elderly demographic

INDUSTRY

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

- All sectors will feel the impact of AI but to differing degrees and in different ways
- In some sectors the alignment of technologies such as AI, robotics, big data, IoT and cloud will enable complete automation of services and processes, leading to the disruption of existing businesses and the creation of new ones
- However even in creative sectors, often viewed as immune to the onslaught of AI, smart machines will help humans make less mistakes, experiment faster, work smarter, manage complexities, see insights and patterns they are currently blind to and so on

7. How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- We need to make it easier for businesses and organisations across all sectors to share their data:
 - We need data marts (along the lines of that created as part of the MK:Smart initiative) to facilitate the exchange of data on both a free of charge and chargeable basis
 - We need standards to facilitate the exchange and aggregation of data but also to enable different datasets to be mashed together to create new datasets
 - We need data passports to hold the metadata associated with datasets and to control what purposes these datasets can/ cannot be used for and by whom

ETHICS

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

- These ethical standards probably need to be entrusted to independent organisations such as the Royal Society
- The decisions that AI models subsequently take will depend on factors such as the way the models have been constructed, their training methodology, selection of the training data and so on. This puts significant responsibility on the humans (or other AIs) who are involved in these processes, not least because of the danger of building bias into these models

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

- Lack of transparency is probably not acceptable in applications where we need to be confident about how the underlying decision process is being driven. One example would be applications that are making decisions about human lives (e.g. judicial sentencing, recruitment) where the risk of bias in the underlying training data is high
- That said, we are clearly going to have to benchmark the performance of AI applications against what humans can actually achieve, including when it comes to the issue of transparency

THE ROLE OF GOVERNMENT

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- Currently there is no disincentive against companies that entice whole departments from universities in areas such as AI, vision systems and robotics. By not leaving behind a critical mass of knowledge and expertise, it takes much longer for the department to regenerate itself and in some cases it never does
- Ethics is an area where the government should be taking a leadership role in terms of deciding who should be entrusted with evolving the ethical guidelines relating to AI, even if it stops short of legislation
- In our experience, negotiating frameworks for IP ownership and exploitation remains the greatest barrier to effective collaboration between industry and universities. This inhibits collaboration in the most exciting "secret sauce" areas and at the moment, AI is definitely in that space. Part of the problem here is the outdated models for measuring research impact, leading to a continuing obsession with publication as the key metric of success

LEARNING FROM OTHERS

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

- China is an interesting example of a society where less legislation around the application of technology is fuelling faster experimentation and innovation, including when it comes to the use of data and AI

- We need to be careful that in attempting to protect ourselves from the future outcomes we might fear, we do not legislate ourselves into a self-fulfilling prophecy that might be just as bad in other ways

ABOUT OCADO

Ocado is the world's largest dedicated online grocery retailer with over 500,000 active customers. We employ over twelve thousand people across 21 sites in England and four development centres in mainland Europe, sell an unrivalled range of products at extremely competitive prices and utilise industry-leading proprietary technology to provide this service to customers' kitchen tables.

Innovation is at the heart of what we do. Using our own technology, we have created a new generation of customer fulfilment centre that uses a hive model in which swarms of robots transport products to pick stations. We make extensive use of AI across our highly automated end-to-end e-commerce, fulfilment and logistics platform in areas such as predictive analytics, monitoring/ oversight and real time optimisation. This technology has a range of applications including: vision systems for robotic picking and decant; processing all the unstructured data in the business, such as emails, voice and social; and processing all the data exhaust from our robots in order to provide analytics and oversight which would be impossible using human engineers watching screens. Furthermore, embedding intelligence into our robots to make them smarter in terms of self-test, exception handling and recovery, as well as facilitating swarm based learning.

Ocado Group plc

1 September 2017

Mr Jeremy O'Connor – Written evidence (AIC0034)

Artificial Intelligence

The ethical implications of artificial intelligence

1. Since ethics are a matter of opinion, I provide an opinion which I have considered and refined over more than 5 years. To keep this essay short, I have minimised external references and the opinions already espoused by respected organisations and individuals.

2. In abstract, I suggest that the inevitable (albeit far-flung) future of artificial intelligence is - Alternative Intelligence. While we may not welcome it, fearing that it might replace us, we need to nurture it because ultimately it will.

3. My argument has two strands:

Artificial intelligence is necessary now

The development of artificial intelligence and the hop across to Alternative Intelligence is an inevitable outcome of human evolution

Artificial intelligence is necessary now

4. Governments and international bodies have recognised that, for at least the sake of humanity, if not for the sake of all terrestrial life, there is a need to expand our foothold in the solar system and even elsewhere in the Galaxy. Whilst humanity can probably deal with Anthropocene calamities such as climate change, pandemics and large scale wars, we may not be able to deal with something like a massive meteoric impact. In a lecture at Cambridge University in 2016, Professor Stephen Hawking expressed concern about the effect of a range of human activities: 'I don't think we will survive another 1,000 years without escaping beyond our fragile planet'.

It has long been recognised that humans are not an ideal life form for existence anywhere but on the dry surface of this planet. Astronauts doing their time on the international space station experience osteoporosis almost immediately on arrival, to name but one item of evidence that in our natural form, we are going nowhere.

5. To fully explore just our solar system, to understand how we can exploit it and to actually exploit it, we need to dwell in hostile environments and work there at scale. We already know that we will not be doing this with humans. But we need to do better than remotely controlling technology that we carefully developed on earth decades before we call upon it to adapt to the unexpected conditions of its mission. Furthermore, current exploration missions are measured in lifetimes of space scientists and mission project directors. Although the benefits of these missions to science are rightly trumpeted, they are meagre given the investment of time and talent. We need to deploy autonomous technology that can take

advantage of local materials and energy elsewhere in the solar system to conduct these missions and indeed to adapt to conditions and even correct the approach to achieving mission goals in a timely manner. Here is the grandest call for artificial intelligence and this theme does not need much more development. We can all imagine and other essays will no doubt expand upon the full uses of it, the work being done to develop it and so forth.

6. The crux of this essay is the catch now raised: although we need artificial intelligence to extend our reach, we are still stuck here on planet earth.

The path to Alternative Intelligence

7. It won't be us that inhabit back-up moons and planets. It literally won't be us. We are all pleased to say that mankind has stood on the surface of the moon - but it was none of us that did that. And yet we project our individual human experience into that event and say 'we did it'.

If, hypothetically, a pioneering group of humans successfully colonises an extraterrestrial environment, then within a relatively short space of time, they will no longer be humans as we know the species. Those that survive, adapt to a changed gravitational force and successfully pass on their genes will eventually undergo allopatric speciation; that is to say, they will become sufficiently different to us by virtue of their alien environment and geographical separation from us that they will effectively become a separate species. One might consider that, after a number of generations, genetic and physiological changes would render a terrestrial human incapable of successfully mating with one of these pioneers to produce a viable child. So according to the definition of 'species', those pioneers will not be Us.

8. So we need to accept that part of our far off future and perhaps the best-case scenario for our survival is that we will generate a new alien species that will live in parallel with us and that potentially will survive us.

9. Now back to technology. If one takes the view that we will never seriously try to inhabit other alien places because we are too squeamish to subject even the most stoic volunteers to the biological challenge outlined above, then for sure we will prefer to substitute artificially intelligent technology for humans in the endeavour. As briefly developed, we can imagine this technology to be autonomous and adaptive. It will be able to repair and reproduce itself, and live off the environment it finds, whether wet, dry, cold, hot or with either extreme of gravitational challenge. It will thrive not only on our moon, Enceladus or Mars, but frankly anywhere it decides to go, since it will not be constrained by human physiology. If it is capable of capturing solar or other energy for fuel on an enduring basis and if it is not much bothered by the passage of time while it mines and refines the materials it needs for its purposes, then it will be free to thrive anywhere. It will have no need of human direction or any other apron strings. We will have laid the ground work for and then watched the development

of an Alternative Intelligence, absolutely distinct from the constrained biological alternative already discussed.

10. By this route we arrive at the ethical issue. We might be right fear that this, by now, Alternative Intelligence (AI) will have no concern for human plight and worse, there is a risk that it may see us as a resource in the same way that we treat absolutely every other living or inert thing with which we come into contact.

11. There are two mitigations to this risk.

The first is that humanity will hopefully not have been complacent over this time and will have developed plenty of human-directed technology and perhaps some improved intelligence to be able to disagree with the new AI if it sees humanity and terrestrial life as a commodity.

The second is a moral perspective that may lend some comfort and which is the essential ethical argument to accept this AI: the AI is the child of humanity. Like our own children, it will not follow the path we imagined or hoped for it, it will challenge us, it will put us in a home - and when we are dead it will move on. Also, like our children, it will not think our thoughts, it's experiences will not be ours. On this basis, why should we ethically reject this outcome any more than we might reject the danger of our children moving life forward without us?

The hop from artificial intelligence to Alternative Intelligence is part of human evolution

12. The argument against the second, moral mitigation above is that our children are our blood and are entitled to inherit our existence, whereas the AI will be an imposter. The counter argument to this is that AI is an inevitable product of human evolution. Let's examine this.

13. We are extremely lucky to have come through natural selection as conscious beings and the positive aspects of society and technology are a testament to our capabilities. It took us something like 3 million years to develop bipedalism and precision grip through our inherited opposable thumbs and the subsequent ability to make our thumb and fingers meet. Our real 'killer app' and true differentiator is the development of incredibly intricate spoken language, a capability that is supposed to have taken around 100,000 to 150,000 years to develop.

Archeology suggests that it took a few thousand years to develop society and for most human groups now, we simply cannot survive for more than a few weeks without that society functioning effectively.

14. If we were to continue to progress along natural, Darwinian lines (and here is a view to suggest we are), we would continue to exploit our natural capabilities and our environment without meaningful constraint, enthusiastically take new risks of death by disease we can't deal with and continue to kill each other for ideological or other irrational reasons. The worst-case prognosis of this commentator is that we would eat ourselves out of house and home, become an

infestation on the planet and eventually fall foul of disease or war. No doubt some sort of furry little mammal would emerge to replace us.

15. Optimistically, with our intellect and our ability to communicate and cooperate, we can take steps to advance rapidly along a novel path of evolution. That is, possibly uniquely, we humans do not have to settle for natural selection as our way forward. We can, for the purposes of survival of knowledge - which is increasingly becoming the most sacred characteristic of humanity - elect to positively evolve our fully adaptive descendant. Through nurturing Alternative Intelligence and accepting it as the beloved child who will succeed us, we may be able to preserve and allow to thrive forever, our human experience of life.

Mr Jeremy O'Connor

30 August 2017

Dr Dan O'Hara, Professor Shaun Lawson, Dr Ben Kirman and Dr Conor Linehan – Written evidence (AIC0127)

Dr. Dan O'Hara¹⁰⁰¹, Prof. Shaun Lawson¹⁰⁰², Dr. Ben Kirman¹⁰⁰³, Dr. Conor
Linehan¹⁰⁰⁴

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1. Anthropocentric misapprehensions about AI are likely to hinder development. Expectations of a generalized human-like intelligence not only conflict with development in other areas of technological advance, but are likely to produce limitations on what benefits can be derived from AI in combination with other technical fields.¹⁰⁰⁵

Current AI is generally encountered as an embodied algorithm in a computational system (e.g. smartphones; self-driving cars; weapons systems) mimicking human behaviour and assisting human activities.

However, developments in robotics have tended away from the anthropocentric and mechanical, mimicking forms of natural intelligence instead. Attempts to produce robots in foam, rubber, and chemical gels, imitating the forms of invertebrates rather than humans, suggest new social and technical applications beyond the normal range of human abilities.

A focus solely on human-based AI is less promising than a focus on other synthetic types of biological intelligence. The potential applications of AI are considerably broadened if AI is taken to encompass a possible future of "bio-hybrid" devices: computer chips grown from bacteria or slime mold, that can be programmed but that can also 'think' for themselves, in the sense that nature 'thinks' its way towards its own solutions.¹⁰⁰⁶

¹⁰⁰¹ Senior Lecturer in English, New College of the Humanities, and co-Director of Virtual Futures

¹⁰⁰² Professor of Social Computing and Head of Department of Computer & Information Sciences, Northumbria University

¹⁰⁰³ Lecturer in Interactive Media, University of York

¹⁰⁰⁴ Lecturer in Applied Psychology, University College Cork

¹⁰⁰⁵ Taylor, A. S., 'Machine Intelligence', *CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), pp. 2109-2118.

¹⁰⁰⁶ 'Growing computer chips from slime mould and bacteria', *Horizon: the EU Research and Innovation Magazine*, 16 Feb 2015

Is the current level of excitement which surrounds artificial intelligence warranted?

2. Insofar as current developments in e.g. machine learning are having and will have widespread and massive social, economic, and military consequences, a certain level of 'excitement' is warranted.

However, excitement about AI, taking the term 'intelligence' literally, is not warranted. Widespread use of misleading terms, loose definitions, and anthropomorphic language such as 'Intelligence' tends to mislead about the nature of the technological advances taking place. There is a great distance between how the potential of these technologies is presented to, and perceived by, the general public and the actual capabilities of AI.

There is no reason to believe that General Artificial Intelligence – the mechanical reproduction of a human-like cognitive ability, self-aware or not – is on the horizon. Dreams of such have been a part of the Western imagination for nearly 3,000 years, and remain unfulfilled. The problem is just as much a semantic as a technical one: if we posit by analogy a field of 'Artificial Emotion', the conceptual cul-de-sac it represents is rather more evident.

How can the general public best be prepared for more widespread use of artificial intelligence?

3. Humanity's outsourcing of decision-making is not confined to AI or even non-AI algorithmic systems. Abrogation or delegation of decision-making capacities is the defining characteristic of all reductive rule-based systems for guiding human action. From this perspective, AI as a cultural phenomenon follows the well-established pattern of belief systems which propose an intelligent and benevolent author of a manual for decision-making containing specific algorithmic rules for exact situations (e.g. the Judaeo-Christian God; the Bible; the Ten Commandments).

The skills and education which the advent of mass AI most demands across the *whole* of society are therefore not restricted to STEM areas, but in fact require a much broader spectrum of historical, philosophical, and social scientific approaches. Investment solely in technological approaches, which are fundamentally ends-driven without necessarily having a clear social end, may result in an AI version of the blind leading the blind.

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

4. At present, those who are willing to grant executive decision-making capacity to AI systems *despite* the absence of true intelligence gain an advantage over human systems owing to sheer speed of operating and reaction time. High-Frequency Trading systems are a prime example.

Their model implies that any field of human activity that involves competition could exploit the capacities of 'dumb' AI, from sports to warfare.

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

5. Efforts should be made not only to increase public understanding of what AI is and is not, but also to make the public less vulnerable to the 'Placebo Effect' of AI. For example, research in games development has demonstrated that telling a player that a game opponent AI is learning and adapting to their behaviour results in the player not only believing in the existence of the AI, but also changing their own behaviour, even if the "AI" is just random.¹⁰⁰⁷ This illusion of complexity shows the need for AI developers to present the exact nature and capabilities of their systems without relying on non-explanations and anthropomorphic associations used today (e.g. 'smart', 'intelligent' systems) that imply humanlike complexity.

See also (3) above.

An additional concern derives from research on the topic of "Human Error" in organisations, which covers everything from major industrial accidents to the misuse of emails. Current thinking has moved away from attributing errors to "bad apple" employees, and towards a view of error as inherent in the design of a system or organisation. There is a drive in management and design research towards reducing error through improving the usability, understandability, and transparency of systems that employees use. There is a genuine concern that the use of AI in the operations of organisations reduces that transparency, affecting the autonomy, power and accountability of employees, and creating new 'grey areas' over the accountability of organisations, with clear legal, ethical, and regulatory implications.

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

¹⁰⁰⁷ Denisova, A. and P. Cairns, 'The Placebo Effect in Digital Games: Phantom Perception of Adaptive Artificial Intelligence', in *ACM CHI Play '15 Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (2015), pp. 23-33.

6. Without commenting on specific sectors, it is worth drawing attention to the unique opportunities the vagueness of AI offers to businesses that may profit from this very vagueness and lack of understanding. As there is no minimum definition of what constitutes AI, it is trivial to claim any system is 'Intelligent'. In the short term 'AI' is liable to be linguistically ubiquitous but seldom actually present in any meaningful sense.

In the longer term, any employment sector that relies heavily on human labour that is easily substituted by algorithms, especially where the work process is already rule-based or involves following flow-charts, is both vulnerable to and liable to benefit from AI (e.g. transport, retail, accountancy, HR). Frey & Osborne's work at the Oxford Martin School is the standard research in this area.¹⁰⁰⁸

However, more caution is needed in generalizing about sectors that differ in practice between cultures and nations. For example, countries with civil law systems may find themselves more vulnerable to AI automation than countries with common law systems that operate by interpretation and precedent, as codified law is more easily automated.

How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

- 7/8. The key ethical issue surrounding all AI systems, from data-gathering monopolies to autonomous weapons systems, is that of accountability. The mythology of AI, perceived by the public as effectively a form of technology equivalent to magic, enables it to be used as a scapegoat in crises – a non-explanation used to evade responsibility for real actions by relying on false perceptions of the autonomy and personhood of AI.

Understanding the distributed levels of interaction with an AI is essential to attributing ethical responsibility. Any single interaction may involve a network of consumer, operator, company, and platform. For example, IBM Watson sells its AI platform services to medical companies for the purposes of diagnosis, but disclaims all ethical responsibility, asserting that the way Watson is used is a matter for the company concerned.

¹⁰⁰⁸ Frey, C. B. and M. A. Osborne, 'The future of employment: How susceptible are jobs to computerisation?', *Technological Forecasting and Social Change* (2017), vol. 114, issue C, 254-280

In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

9. Any degree of lack of transparency or 'black boxing' whatsoever in AI systems that have capacity to act is only morally equivalent to the same, inevitable lack of transparency in the accumulated knowledge and decision-making power of an expert human. Traditional safeguards against abuse of such power consist of such commitments as the Hippocratic Oath, and rely both upon the identifiable agency of the expert, and the existence of equally-expert peers to judge observance or breach.

It would be a mistake to anthropomorphize the agency of an AI system, and it may be impossible to judge with equal expertise.

Hence the most effective safeguards are likely to be those that clearly attribute human responsibility in advance – however unfair this may seem – and those that prevent the end-user from over-disclosure of data. Both are safeguards that can and should be integrated into current design processes.

In relation to transparency, Baroness Kramer has repeatedly asked, most recently at the APPG on the 10 July 2017, about the potential hazards of compounded error embedded in AI systems. Could a small initial error, reproduced and magnified by a machine learning system, produce unforeseen and uncontrollable effects – a runaway iteration, as it were?

There is a considerable amount of fundamental research upon the potential problems of non-human agency and compounded error, from which AI development could and should learn. Concepts of path-dependence and skeuomorphism outline the ways in which errors in initial design or misapprehensions in reproduction can become embedded in a repeated design process over thousands of years.¹⁰⁰⁹

In an AI example, Microsoft's repeated experiments in AI conversation systems were trained by 'trolls' to use racist language, which demonstrates a clear failure in designing for compound effects. Similar design failures in less transparent systems have the potential for serious consequences.¹⁰¹⁰

Where there is hazard of serious malign effects, existing ways of regulating and safeguarding complex dynamical systems such as Air Traffic Control should provide a model for managing and mitigating such effects. It is however at the design stage that AI developers such as those working on neural networks can best address the unpredictability and opacity of the

¹⁰⁰⁹ O'Hara, D., 'Skeuomorphology and Quotation', *Morphomata* 2 (2012), 283-94

¹⁰¹⁰ Singh, I., Walden, I., Crowcroft, J., and J. Bacon, 'Responsibility & Machine Learning: Part of a Process' (October 27, 2016). Available at SSRN: <https://ssrn.com/abstract=2860048>

system in question,¹⁰¹¹ and those working on machine learning systems can best address the encoded biases of a system.¹⁰¹²

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

10. AI should be subject to government regulation for the very reason that it is *not* intelligent. It therefore differs only in degree, not in kind, from existing activities that are subject to regulation.

Policy-makers could pay special attention to the possibility of recursive effects of AI upon existing ways of conducting human activities. For example, if assistive AI systems are adopted broadly in decision-making in the law or medicine or corporate governance, they are likely not only to assist but also to change fundamentally the processes of law or diagnosis or governance as we currently understand them.

6 September 2017

¹⁰¹¹ Holmquist, L. E., 'Intelligence on tap: artificial intelligence as a new design material', *interactions* 24.4 (June 2017), 28-33

¹⁰¹² Caliskan, A., Bryson, J. J., and A. Narayanan, 'Semantics derived automatically from language corpora contain human-like biases', *Science* (2017), vol. 356, issue 6334, 183-186

Dr Seán Ó hÉigeartaigh, Dr Shahar Avin, Martina Kunz, Andrew Ware and Dr Simon Beard – Written evidence (AIC0150)

**Dr Seán Ó hÉigeartaigh, Dr Shahar Avin, Martina Kunz,
Andrew Ware and Dr Simon Beard – Written evidence
(AIC0150)**

[Submission to be found under Dr Simon Beard](#)

Dr James O'Shea – Written evidence (AIC0226)

0. Introduction

- 0.0 I am currently a Senior Lecturer at Manchester Metropolitan University and a director of an AI startup company, Silent Talker Ltd. The opinions expressed here are entirely my own, but are based on my research, teaching and technology transfer experience. This includes DTI funded consultancy to industry during a similar period of technological change, the adoption of microelectronics during the early 1980s.
- 0.1 I am an investigator on the Horizon2020 iBorderCtrl project which has 13 partners across Europe (including 4 border agencies). This project is intended to speed up the crossing of 3rd party nationals and freight (e.g. the UK post-Brexit) into the Schengen Area. It uses a patented (Rothwell et al., 2002) AI based deception detection system (Silent Talker) developed by myself and colleagues, in a pre-travel interview. This is combined with other measures (Biometric, Document etc.) to assess risk and guide border guards in their dealings with travellers. The pre-travel interview uses questions nominated by actual border guards. Silent Talker is a specific application of a generic technique (adaptive psychological profiling) developed in my research group to assess a person's internal mental state through non-verbal behaviour.
- 0.2 Another aspect of my research is Conversational Agents, in which a computer uses natural language conversation to lead a human through a complex application, such as advice on debt management or dealing with workplace bullying and harassment (Crockett et al., 2009).
- 0.3 I was co-founder of the MMU Intelligent Systems Group in 1992 (AI research). I was chair of the 2011 Agent and Multi-Agent Systems, Technologies and Applications conference (an important branch of AI). I have taught Software Engineering and Artificial Intelligence to final year undergraduates, with a strong emphasis on employability, for over 20 years.

Rothwell J., Bandar Z., O'Shea J., McLean D. Analysis of the Behaviour of a subject. UK Intellectual Property Office, WO 02/087443 A1. 2002.

1. What is the current state of artificial intelligence and what factors have contributed to this?
 - 1.1. Traditionally Artificial Intelligence has been divided into two camps, those who want to build systems to perform tasks which require intelligence when performed by humans and those who want to build systems indistinguishable from humans in terms of consciousness, felt emotions etc.
 - 1.2. The first camp has achieved considerable success with tasks which are too complex to define for traditional computer programming ("coding") but have limited scope and for which we can obtain example

cases to train a specialised AI system to solve the problem. For example, we can train specialized systems for medical diagnosis and to decide whether to approve a mortgage – but we not expect the mortgage system to diagnose diabetes.

- 1.3. The second camp has made progress in modelling consciousness, but we still seem to be a long way from developing a conscious machine (Reggia 2013). In particular, creating a machine which exhibits the genuine feeling of consciousness seems particularly intractable (Lisman, 2017). These points are important when dealing with public perception of AI.
- 1.4. There are interesting possibilities for AI at the interface between the two camps. We are developing systems, which simulate elements of consciousness or emotion, along with more routine AI components to solve complex problems. For example, my group's contribution to the Horizon 2020 iBorderCtrl project in which we are developing an Avatar (computer generated artificial person) who will present neutral, positive/encouraging or puzzled/skeptical emotions to the interviewee, depending on the degree of deception detected in answering pre-travel questions.

Reggia, J.A., 2013. The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks*, 44, pp.112-131.

Lisman, J.E., 2017. Locke's View of the Hard Problem of Consciousness and Its Implications for Neuroscience and Computer Science. *Frontiers in psychology*, 8

How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

- 1.5. AI has a track record of technological leaps followed by periods of stable, steady progress or stagnation, for example the "AI Winter" in the 1970s following the publication of "Perceptrons" by Minsky and Papert (1969).
- 1.6. We appear to be moving from a stable/positive period to another leap due to a combination of Big Data (large collections of data which we have been able to amass and access over the last couple of decades due to advances in mainstream computing) and new specialized hardware designed specifically to exploit this with AI (<http://www.nanalyze.com/2017/05/12-ai-hardware-startups-new-ai-chips/>).
- 1.7. So I expect an increase in AI capability over the next 5 years, a plateau at around the 10-year mark, then a period of consolidating benefits leading to a possible further step up in capabilities around the 20-year mark.
For the current step to succeed, the new accelerator hardware has to be effective, big data has to be successfully collected, cleansed & balanced and there must be sufficient skilled scientists in the workforce to

implement this. Past experience suggests the UK should focus on creation of big datasets, educating the next generation of AI scientists and commercial software development.

2. Is the current level of excitement which surrounds artificial intelligence warranted?
 - 2.1. Public perception of AI is subject to contextual influences, so excitement is warranted, but it is often the wrong kind of excitement.
 - 2.2. The public are virtually oblivious to established / mainstream applications of AI which improve their everyday lives. They take no particular interest in whether or not AI underpins recognition of their number plates or the decision to give them a mortgage.
 - 2.3. The public are prepared to confer human or superior capabilities AI when it presented in a particular way. Many speak of their Satnav systems as if they are human ("she's always angry with me"). Some have reacted to the animatronic artwork of Jordan Wolfson (https://youtu.be/8ppmjP_U9mw) as if it is a genuine woman, but it is closer to a puppet than a robot (Penny, 2016). The public have yet to learn that AI is more than skin deep.
 - 2.4. The "Terminator" myth. Hollywood has promoted the idea of AI equaling or surpassing human intelligence, leading to an apocalyptic war between humans and AI. My view is that this is not likely in the foreseeable future because to engage in a genuine war requires strongly felt emotions and other traits of consciousness.

Penny, S., 2016. Robotics and Art, Computationalism and Embodiment. In *Robots and Art* (pp. 47-65). Springer Singapore.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

- 3.1. Jobs: There will be continued displacement of successively more skilled people from the workforce. On the other hand, there will be and the creation of a (relatively small) number of high skilled jobs in development and support of AI applications.
- 3.2. Education is an important key to the future success of the UK in the AI market. AI degrees will require highly developed STEM skills.

- 3.3. This is highly challenging to students and universities find it a disincentive to promote AI units or full degrees as they may lead to lower NSS scores. Some policy changes may be needed to counter this.
- 3.4. Many of the best-paid AI existing jobs require a PhD in an AI topic. We should be more aware of the potential for a computer scientist to graduate, work for 5 – 10 years in industry, then return to education to do a PhD. Taking a 3-year career break does not fit into our current career models but undertaking PhD part-time, mid-career, entails a high risk of failure.
- 3.5. It is never too early to start educating people for AI careers. At its most rewarding, working in AI is a form of creative play for adults. Therefore, we shouldn't see preparation in purely terms of drilling people to do "hard sums" (teaching algebra, logic, calculus etc.). Early learning (from the nursery onwards) should also encourage creativity, problem solving, verbal fluency and crossing disciplinary boundaries. Education in the humanities, including philosophy, will contribute to producing good AI practitioners and good citizens who interact with AI.
- 3.6. The most robust non-AI jobs will be those relying significantly on aspects of consciousness such as style (artworks, fashion, design, crafts etc.) and empathic emotion such as caring, nursing etc.
- 3.7. AI has a lot of scope to provide "physical" interventions to support aging and disabled members of society in independent living. such as predicting falls for the elderly, acting as "broker" agents for people with communicative difficulties and in learning support through intelligent tutoring for pupils with learning difficulties (one of my current PhD projects concerns science education for school pupils with high-functioning autism). There is some potential for job creation here to exploit a synergy of human and machine – Robots may be good for lifting people in and out of bed, humans may be better at motivating people to get in and out of bed.
- 3.8. Other positive societal effects include: adaptive personal tutoring in education, security, crime detection/prevention and counter-terrorism, assisting the general population with more complex tasks in their lives. Humans will continue to work in these fields, but may need education about working with AI to perform their jobs.
- 3.9. Negative effects will include the potential loss of civil liberties due to increased efficiency of mining personal data and monitoring people. Solutions could take different forms. We could counter the abilities of AI with more pro-active, intrusive and statutory personal data protection. Alternatively, we could make a cultural switch as a society to the position where more is known but less is cared about our personal lives.
- 3.10. A question I have been asked on TV and radio interviews is "Suppose your AI lie detection technology becomes generally available, what will happen in a society where no-one is able to lie?" If there are no more social white lies, bluffing and haggling in negotiations, face saving excuses etc. it is possible that we may have to radically change our

society's *mores*, to gloss over the unpalatable truths that AI may reveal to us about each other, simply to continue to function.

- 3.11. This prepares the ground for my recommendation that the committee should also seek evidence from philosophers – as an AI researcher I have found their work to be productively informing in the past.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

- 5.1. Public perception of AI is subject to contextual differences.
- 5.2. The public are virtually oblivious to established / mainstream applications of AI. They take no particular interest in whether or not AI underpins recognition of their number plates or the decision to give them a mortgage.
- 5.3. The public is prepared to confer human or superior capabilities AI when it presented in a particular way. Many speak of their Satnav systems as if they are human ("she's always angry with me"). Some have reacted to the animatronic artwork of Jordan Wolfson (https://youtu.be/8ppmjP_U9mw) as if it is a genuine woman, but it is closer to a puppet than a robot (Penny, 2016). The public have yet to learn that AI is more than skin deep.
- 5.4. The "Terminator" myth. Hollywood has promoted the idea of AI equaling or surpassing human intelligence, leading to an apocalyptic war between humans and AI. My view is that this is not likely in the foreseeable future because to engage in a genuine war requires strongly felt emotions and other traits of consciousness.
- 5.5. We are far more likely to fall foul of "cock-up" rather than conspiracy. They may be future occasions in which human lives are lost unintentionally through malfunctioning weapons systems, but this will originate in human errors in constructing the AI components or operating them on the field of battle.
- 5.6. My experience as a science communicator is that media interviewers will always throw in a couple of "sexy" questions to thrill their audiences. Scientists are naturally tempted to play along with such questions because they need to sell their science to create perceived impact to support funding. Part of the training for young scientists must be to tread the line between excitement and sensationalism carefully.
- 5.7. We have many potential opportunities to engage with the public though festivals, science fairs etc. in which we have more control over our messages than through the mass media. However, there are often

obstacles in the way that could be surmounted with *very easy* access to *very small* pots of money for things like travel expenses, consumable materials etc. Government could facilitate this.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

5.8. Security, Criminal Investigation and Defense. By way of example, my research group's Horizon 2020 project (iBorderCtrl) uses AI to speed up the crossing of 3rd party nationals and freight (e.g. the UK post-Brexit) into the Schengen Area without compromising on security.

AI may also be deployed to defend society from IED attacks, firing into crowds (and other attacks on crowds), detecting and preventing radicalization leading to attacks on civil society and prevention of criminals and terrorists from entering the UK.

Healthcare and related industries, see 3.7

Practically every sector can benefit in some way from AI, even though it may be indirectly. Even if there is no obvious application for AI, if the sector requires generic activities such as scheduling and planning, AI can contribute.

However, I suggest that the people least likely to benefit would be those relying significantly on aspects of consciousness such as style (artworks, fashion, design etc.). and empathic emotion (human aspects of caring, nursing etc.). (i.e. their jobs will be safe because AI cannot contribute as effectively to doing those jobs).

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

On the one hand, large corporations will put significant (financial) investment into collecting (and hopefully vetting) big data collections and will wish to protect their investment while they produce saleable AI from it.

On the other hand, I would be unhappy with (for example) large government IT projects depending on the fruits of such data if it were not available for independent scrutiny. Therefore, some disclosure requirement should be included in Government procurement contracts

I would also be unhappy for companies to have a monopoly on data extracted from government or public bodies as the original collecting source (e.g. census and NHS data).

I can see possible circumstances under which monopolies would be highly anti-competitive (e.g. an insurance company gaining and underwriting advantage through better knowledge of health risks). There could be a requirement for

some (high level) transparency when a large organisation has markedly different pricing / policies from the market as a whole.

I do not believe you should be able to patent data, but maybe there could be some form of copyright with a much shorter protection period than (say) a literary work.

If a company wishes to patent a method, process or other invention derived from a large dataset, the relevant data should be published as part of the patent application (sufficient information for one experienced in the field to reproduce the invention).

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

- 8.1. The most important aspect of this is the degree to which AI systems are granted autonomy. An extreme example would be fully autonomous battlefield drones. According to the EU report "Human Rights Implications of the Usage of Drones and Unmanned Robots in Warfare", the technology to make fully autonomous drones does not exist at this time. ([http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/410220/EXPO-DROI_ET\(2013\)410220_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/410220/EXPO-DROI_ET(2013)410220_EN.pdf)). The report does describe systems such as BAE Systems Taranis intended to incorporate autonomous capabilities with a human in overall control. This is known as "human-in-the-loop."
- 8.2. There are also human rights issues in high-stakes non-lethal applications of AI such as those leading to a citizen being arrested, denied employment, denied entry to a country or dismissed as a victim of an alleged offence.
- 8.3. In my opinion these issues hinge upon whether an AI system takes a decision about a person or simply provides evidence to a person who makes a decision – a return the human-in-the-loop principle.
- 8.4. The right to dignity. I do not see much of an issue with this provided the interfaces to AI systems whose interfaces are designed sensitively. AI components do not have emotional / subjective properties. The only potential issue I can see is one of bias introduced during the production of AI components. This is covered in 8.6
- 8.5. Racial or cultural discrimination. Lack of consciousness and emotion works in favour of AI systems in their claim to robustness against discrimination. However, human developers need to take care not to build bias in at the training stage. This could be a particular weakness where Big Data is used and the data is not properly vetted, cleansed and balanced. Suppose I built a lie detection system trained from "white Europeans" and then operated the system on another ethnic group? If the second ethnic group had different cultural patterns of non-verbal behaviour, they might produce non-verbal indicators of deception when telling the truth and be unfairly labelled as liars. My current stance is that, in my work, specific AI classifiers should be developed for each ethnic group using the system, so that each is treated fairly. Part of the

iBorderCtrl project involves testing whether such differences exist and if so at what level of granularity is ethnic / cultural partitioning most effective in producing fair classifications.

- 8.6. In AI research, different universities have their own ethics procedures. Some of these require tedious and repetitive filling of the same sets of forms for each minor variation of an AI experiment.
- 8.7. Many research projects require international partners and the satisfaction different ethical standards across different countries.

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

Using my experience with the iBorderCtrl project, described in paragraph 0.1 to frame an example, I see some complications. These arise from legislation (EU Directive 680/2016/EU, which I presume will have some equivalent in UK law for the foreseeable future): *"subjects of biometric decision making have a right to be informed of automatic decision making, that it is made transparent and that the subject has the right to express his/her point of view or the right to contest the decision."* These types of question need to be considered by your committee.

At what point does an AI system take a decision about a human instead of simply collecting evidence?

What responsibility is there on the AI system or its developers to explain how the system has reached its decision (we use a form of AI called an Artificial Neural Network which is a black box and effectively inexplicable to humans. We have experimented with producing a rule-based equivalent that produces over a thousand rules using collections of logic operators – again unsuitable for the average traveller).

What is the equivalent duty for a human? If a border guard suspects an interviewee of deception, what is the responsibility and degree to which the human must justify / explain the decision (human explanations of how they reach decisions may be instinctive and have no effective predictive explanation)?

Should you tell the traveller during a pre-travel interview if he / she is suspected of deception? If so, what is the mechanism for contesting the "decision"?

Should you answer questions on how to pass and so lead the traveller to believe that he / she is getting coaching on how to appear truthful?

My view is that "inconsequential" decisions by AI components (i.e. the traveller was truthful, no action needed) do not need to be explained to travellers or contested by them. Where a traveller is suspected of deception, the AI system should provide evidence to a human-in-the-loop, who will take the decision and comply with the traveller's rights.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

With reference to 7.8, I strongly suggest that the government level the playing field of the bureaucratic workload by commissioning a group of experts to study the top 10 Russell group university ethics procedures and from them synthesise a national ethics process, establishing the minimum bureaucratic requirement to satisfy ethical needs – so that exactly the same forms, in the same numbers are used in all UK universities. I challenge any academic to produce a justification for a special exceptional ethical process in any particular university (although it may be that different forms are required across disciplines, medicine being different from computer science).

Point 10.1 may seem minor, but it is a serious drain on human resources that would be better spent on active research.

There is a serious gap in support for commercialisation in the UK. UK investors are not prepared to invest in getting a technology out of the laboratory and into its early commercially exploitable stage – this puts us at a serious disadvantage compared to the USA where investors are more proactive.

This is also true of defense procurement. Having applied to the Centre for Defense Enterprise in the past and scrutinized successful case studies of applications, I have realised that CDE can only really support the migration into defense of what is basically a mature technology in some other market sector. This makes defense dependent on serendipity in the civilian AI market.

I do not think the solution to points 10.3 and 10.4 is simply for the government to pump funding into startups, but I do believe that government has a role in changing UK investment culture. This might be done in the case of 10.3 through some form of tax breaks for investors at the higher-risk early stages of commercialisation; however, the process would need to be designed to reward success more than failure.

Education is an important key to the future success of the UK in the AI market. However, AI degrees are challenging and universities find this a disincentive to promote AI units or full degrees as they may lead to lower NSS scores. Some government input may be needed to counter this.

Many of the best-paid AI jobs require a PhD in an AI topic. The solution may be government promotion of career sabbaticals or part-time PhDs for IT professionals. Again this may be supported through tax breaks. I have significant experience of the part-time PhD route and it is very challenging with a high failure rate at present.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

With reference to 8.9, work with other countries to simplify the process of getting international ethical approvals.

To continue to support our participation in EU Horizon and future EU programmes vigorously. The government should try to keep our level of participation and acceptance at the same level as it is currently.

17 October 2017

Alex Olson – Written evidence (AIC0002)

I am a 3rd Year undergraduate studying Artificial Intelligence at the University of Edinburgh.

In summary: We are at the beginning of a technological revolution which will do to middle class jobs what the industrial revolution did to farmers and manufacturing workers.

How can the general public best be prepared for more widespread use of artificial intelligence?

I believe that the largest risk to the average person from Artificial Intelligence is that more and more jobs will become automated as the technology improves. This should not be understated – there is a significant incentive for a company to replace a human worker with a computer, and the technology is improving steadily enough for many jobs to be at risk in the next decade or so. The government should understand that we are beginning to move into a completely unprecedented era of human development, as sensational as that may sound, where it may simply not be necessary for every person to be in work.

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Currently the benefits are only being reaped by corporations – those such as Google who develop new systems, and those such as Tesco who put them in place (e.g. self-checkouts, which have replaced human staff at scale). This is especially damaging for the working class, who have a diminishing pool of jobs to even apply for. However, the effects will not be limited to the working class – this will begin to affect middle class jobs in the next 5 to 10 years, and in some places is already doing so. For example, insurance adjusters in Japan are being replaced by computer systems which are capable of rapidly processing claims and determining their validity. The government must ensure that when this type of automation happens, there is sufficient support available for those who simply don't have alternative jobs to apply to – such as those who might have worked in an industry for their entire lifetime, have skills and training, but in a field which simply no longer exists for humans. It is critical that they are not blamed for their unemployment, and instead helped to transition into what could be a completely different area of the workforce. In the long term, it will not even be possible to transition into a different area of the workforce for many, as there will be just as few jobs in other areas.

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

The only jobs that are not at *immediate* risk of automation are those which require a high degree of creativity – artists, high-level scientists, politicians. Any

job which follows a routine, a set of rules, can and *will* be automated. We must accept that this will happen, because it *will* happen, and begin to prepare for rising unemployment across the working and middle class.

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

I don't believe that artificial intelligence can be regulated effectively due to the pace of advancement. The government's role needs to instead be to ensure that the benefits of artificial intelligence can be enjoyed by all. After all, it is a good thing if menial jobs are automated, but only if the people who used to do them are still supported.

In conclusion, technology is going to continue to rapidly advance, and jobs are going to be automated faster and faster. It won't be long until we are facing an entirely new type of economy, in which there literally aren't enough jobs for everyone. The government needs to accept that this 'decoupling' of work and living standards is inevitable, and not a part of any party ideology. It would happen under Labour, it would happen under the Conservatives. Instead, the government must understand that this is the direction that society across the world is headed, and begin to develop a British approach to this 'post-work' economy.

24 July 2017

Onfido – Written evidence (AIC0163)

Thank you for the opportunity to submit evidence to the House of Lords Select Committee on AI.

Onfido builds trust in an online world by helping businesses digitally verify people's identities. Using Machine Learning technology (a subset of AI), Onfido validates a user's identity document and compares it with their facial identifiers. An innovator in the Computer Vision space, Onfido's machine learning technology learns to identify fraud as it evolves over time, enabling clients to rapidly onboard more users while protecting themselves against fraudulent activity.

In our response, we are defining AI as a computer system able to make and/or learn how to make complex decisions.

Onfido's response to the call for evidence is below.

Impact on society

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

One significant benefit being driven by the implementation of AI is the banking of the 2.5bn unbanked people worldwide.

Historically, there has existed a trade-off between access and security. On the one hand, 2.5bn people in the world are under or unbanked and therefore cannot access financial services. On the other, identity fraudsters are using financial services to launder money used in human trafficking, drug trafficking, terrorist financing etc. This amounts to around 2-5% of world GDP, approximately equal to \$2tn.

By using Machine Learning, Onfido is one of the companies beginning to bridge the gap and reduce the access-security trade-off. Onfido is able to reduce risk such that FinTechs and banks are able to bring thin or no- credit file individuals onto their platforms. This grants access to the under- and unbanked online, whilst at the same time, minimises the chances of identity fraudsters laundering money. This has helped online banks provide debit accounts to any of the 4m unbanked in the UK.

There are also those that stand to be negatively impacted by the development of AI however, and the potential for job losses is a particular concern. In the case of Onfido's technology, Machine Learning is not intended to replace so much as augment human compliance roles. By automating up to 95% of typical Identity

Verification cases, expert human resources are able to give more focus to the remaining 5% that require human intervention.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

Financial services is one area that is already seeing considerable benefit from the use of AI.

Incorporating AI into financial services can lead to reduced fraud, lower operational costs and the automation of compliance processes. This can range from AI-enabled data collection, to analysis and risk-modelling that helps businesses take advantage of that data.

At Onfido, we use prediction and automation techniques to help automatically classify, extract data from and verify identity documents that are provided to us as part of KYC and AML checks.

There is still considerable resistance to the implementation of AI in some areas of financial services, however. While AI solutions are already being used by more agile online banks, it's seeing slow adoption from more risk-averse, highly regulated, traditional financial institutions.

With competitive pressure increasing, financial services are starting to actively seek out solutions to their pain points – but without the infrastructure in place at a high level to support this, it is difficult for incumbent financial services to embrace innovation without exposing themselves to risk. Initiatives like industry sandboxes (of which Onfido is a part) can be really helpful in encouraging collaborative innovation in a safe space and will hopefully see further uptake and development of AI solutions.

Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

The Government could have the greatest impact on AI innovation in the short term by opening up access to EU and global talent pools. Unfortunately, the UK does not have enough homegrown talent to drive the development of AI, and will struggle to keep up without greater access to overseas specialists. Longer term,

more focus on science and technology education would ensure that the necessary skills are being developed to build the workforce of the future.

The development of AI, particularly Machine Learning, commonly requires the use of a significant amount of data, including regulated personal data. We encourage the Government to address the tension between data protection, and the development and use of AI in the forthcoming Data Protection Bill. We would like the Government to make clear that processing personal data for the purpose of training data sets is in the legitimate interest of the controller, and does not amount to profiling that might produce legal effects or significantly affect the individual.

6 September 2017

Online Dating Association – Written evidence (AIC0110)

The Online Dating Association (ODA) was established in 2013 as a trade body to represent the views and interest of the rapidly growing digital dating industry. Alongside its activities representing industry interests, it works closely with stakeholders providing consumer safety advice and channels of communication with law enforcement and NGO's. Central to its remit is the provision of a safe user experience led by its members' adherence to the globally leading ODA Code of Practice.

Online dating accounts for 30% or more of new relationships and there are over seven million registered users in the UK.

Responses to this consultation are done so from the perspective of online (digital) dating services and may not represent the views of all ODA members.

onlinedatingassociation.org.uk

Questions

The pace of technological change

- I. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?
 - a We are at a particularly exciting stage due to the commoditisation of AI, making it more accessible to smaller businesses, either through plugging into managed AI services from Google, Amazon, Microsoft etc, or accessing business which use AI to deliver a service. For example, ODA member Scamalytics, uses AI to catch scammers on dating sites. For many businesses AI services will be provided by trusted third parties. The key benefit being that they can concentrate on making AI better for their customers.
 - b Technology penetration amongst consumers and availability of data has accelerated the development of AI and the growth of applications for AI. It is envisioned that this pace of change will only accelerate as its opportunities are explored further.
2. Is the current level of excitement which surrounds artificial intelligence warranted?
 - a Yes, but this should also be approached with caution. AI is being used as a 'catch all' term for a lot of applications that collect, analyse and produce responses from data sources. Marketing and media activities

- often confuse AI with other technologies, either to sensationalise a story or tap into the current fervour around developments in this area.
- b Coupled with machine learning, AI presents opportunities for improved services to consumers and wider civic society; perhaps personalised offerings and communications.

Impact on society

- 3. How can the general public best be prepared for more widespread use of artificial intelligence?
 - a Society is already experiencing a seismic shift due to the ready availability of technology; in their hands, homes and work places. Key benefits include mass processing of data that enables these services to combat fraud, reduce cyber threats to customers and in the context of dating, provide a safer environment with better matches and more tools and insight to combat scammers and predatory behaviours. AI based systems can also deliver content to users more accurately allowing for the disruption of business models; cost-effective advertising and clearer benefit delivery to users, improving competitiveness of markets. Users of dating services may also find AI useful when looking for a match or creating their own profile.
- 4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?
 - a Broadly, the introduction of AI will lead to retraining requirements as certain jobs are replaced by the technology whilst others are created. For example, certain customer service enquiries can be better handled by AI; which may even learn which responses to similar questions get the best results. However, other roles will be created where the output from AI systems will need a human 'sense check'. Criminals will certainly find life harder as AI can dynamical change security systems to combat new methods of penetration; these will include scammers try to 'con' money out of other users; hackers attempting to access customer and company data and financial fraud being attempted against a company.

Public perception

- 5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?
 - a As a term, AI can be difficult for non-technical users to understand what is meant. However, the outputs of AI systems are much easier to understand. For example, better matches with possible partners;

targeted advertising / content and risk mitigation in the online environment.

- b Engagement and understanding are better sourced by industry as clear messaging helps increase engagement and more efficient business. Public education via NGO's and trade associations can help build trust and keep pace with innovation. The danger of a centralised Government approach is that guidance can be out of date very quickly.

Industry

- 6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?
 - a Online dating services can benefit greatly from the introduction of AI based services. Users see the benefits in terms of improved matches; new business models; protection from predatory behaviour and scammers and innovation leading to creative ways in which to engage with the platform, service or users.
- 7. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?
 - a The introduction of GDPR and other associated consumer protections provide the basis for data sharing. Commercial imperatives are already seeing many of the big service providers opening up elements of their AI services to smaller entities and research facilities. In an increasingly data driven global economy it is important to ensure that frameworks are in place to allow for an increasingly border-less consumer is able to access services globally and for international businesses of all sizes to service these customers.
 - b Data held by Government, available with appropriate protections, can also be made available in an 'open' environment to help consumers access the most suitable services and for business to protect them from cyber security threats and fraud.

Ethics

- 8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?
 - a AI in dating relies on users sharing both personally identifiable data, and sensitive data. For example, if a scammer uses a certain email address, users would expect us as an industry to block that email address from future registrations, even though it is personally identifiable. Clearly for this to work, all users must consent to data sharing, not just the scammers! AI takes this to another level, as it can

- learn characteristics such as conversational style, and use that as a weighting in fraud detection.
- b Matching of users requires consideration of sensitive data such as their sexuality or ethnicity. Generally speaking, AI is ethically "blind" so it is important that humans are involved to ensure that AI doesn't, for example, in the performance of "blind" predictions, start stereotyping.
 - c Government should be wary of imposing ethical requirements which might hamper the ability of AI to prevent fraud or match people according to their religious or ethnic preferences (something quite unique to dating).

The role of the Government

9. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?
- a As previously mentioned, the pace of change can make the regulation of technologies, such as AI, very difficult. However, the outputs can be more clearly covered. For example, we already have strong data protection regulations in the UK and these are being bolstered by the imminent introduction of GDPR. Likewise, consumer law provides protections around contracts, unfair behaviours, advertising, payments and other critical areas. In the dating business context, these regulations are generally output focused and any business activities coming out of AI would be covered here.
 - b The challenge for Government probably lies in the fact that different industries and services don't operate in these frameworks and we would urge that careful consideration is given when looking at a 'high risk' AI environment and unintended consequences that impact other industries, such as online dating.

5 September 2017

ORBIT The Observatory for Responsible Research and Innovation in ICT – Written evidence (AIC0109)

Authors

Bernd Stahl, De Montfort University, Centre for Computing and Social Responsibility

Marina Jirotko, University of Oxford, Department of Computer Science

Martin de Heaver, De Montfort University Centre for Computing and Social Responsibility

What is ORBIT?

ORBIT, the Observatory for Responsible Research and Innovation in ICT is a project funded by the UK Engineering and Physical Sciences Research Council (EPSRC). The purpose of ORBIT is to foster and disseminate a culture and climate of responsible Research and Innovation across the ICT research community. As part of its activities, ORBIT has been active in the ad hoc advisory group for the APPG AI.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

The current success of AI is down to a combination of machine learning algorithms, high performance and widely available computing hardware and the availability of large amounts of data. Large increases or improvements in any of these factors may lead to accelerated development of AI. Such developments may arise from technologies that are currently approaching maturity, such as neuromorphic computing or quantum computing. Further developments of big data generation and analytics, for example created by more widespread use of automatic sensors can have the similar effect.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

AI has had some very high profile recent successes which range from winning the world championship in the game Go to successful proofs of concept of automated cars. However, at this point it is not clear that this translates into progress in terms of general intelligence which does not require specific training in a clearly defined subject area. Generalised intelligence that can draw inferences from limited amounts of data, be situated in a specific context and can be run on low levels of energy would be another game changer but is currently not recognisable on the horizon.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

The general public needs to develop critical reflective capabilities in order to engage with AI in an appropriate manner. This means that a broader engagement with computers and computer science in primary, secondary and tertiary education is desirable. In addition, educational offerings will be required for people of working and retirement age.

Computer literacy and critical thinking skills with regards to computing technologies and media content will increasingly become general life skills that every member of modern societies will need to have. Citizens will need to understand how modern technologies can be used to manipulate them economically as well as politically in order to be able to engage in a rational discourse about the desirable shape of technology and its regulation.

A particular point of concern will be those individuals whose employment will be negatively affected by AI. These individuals will need training and reskilling in order to be able to productively contribute to society.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

The majority of advantages are related to businesses and profits. AI technology can increase productivity and streamline business processes. Given the importance of access to large datasets for current machine learning algorithms, it is likely that the largest technology companies are in the best position to benefit from AI. In order to avoid a monopoly or oligopoly situation, it would be important to ensure that access to training data is widely available. In addition it is likely that redistribution will be required in order to ease the impact on employment. This is likely to involve some level of taxation as well as an investment into training and capacity building.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

A broader societal discourse concerning the way technology is developed and used in modern society is highly desirable and probably required if serious repercussions are to be avoided. Artificial intelligence would figure prominently in such a broader public engagement initiative. It would be in the interest of both large technology companies and policy makers to start this type of public engagement exercise soon. The AI industry has already started to develop structures for this purpose. The Partnership on AI, for example, seems well positioned to represent industry in such endeavours.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Not responded to

7. How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Not responded to

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

There is a significant amount of research that has been undertaken around the ethical issues of AI. Examples of such issues includes threats to privacy, biased decision making, killing of humans by autonomous weapons and many others. This is the key area of activities and interests of ORBIT. Instead of listing issues, we would like to propose that the enquiry need to find a **method** to collect and evaluate issues and translate shared concerns into practical action.

We propose that the concepts and practice of responsible Research and Innovation (RRI) offer such a methodology. The approach to RRI adopted by the Engineering and Physical Sciences Research Council (EPSRC) focuses on four aspects which are reflected in the acronym AREA:

- 1 **Anticipate** – describing and analysing the impacts that might arise.
- 2 **Reflect** – reflecting on the purposes of, motivations for and potential implications of the research.
- 3 **Engage** – opening up such visions, impacts and questioning to broader deliberation, dialogue, engagement.
- 4 **Act** – using these processes to influence the direction and trajectory of the research and innovation process itself.

RRI aims to ensure that the processes and products of research and technology development are societally acceptable, desirable and sustainable.

RRI is a process that allows the continuous reflection of research and innovation and its products. It is important to underline that dealing with ethical and human rights implications of **AI will require a process**, not a one-off solution. The technology and the issues it raises will continue to evolve and today's solutions are unlikely to be applicable tomorrow. RRI can structure this continuous engagement. The components of anticipation, reflection and engagement with stakeholders and the broader public need to be institutionalised to have an ongoing way of dealing with them.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

Most advanced technologies are not completely transparent to most users most of the time. This is typically accepted in light of the advantages of the technology brings. The limit of acceptability is reached where technical artefacts have significant impact on human beings that require an ability to scrutinise them and, where possible rectify them. Examples would include safety-related questions as in autonomous vehicles or automated decisions concerning credit rating, employment options, education, housing etc. based AI-based systems. In such cases a right to review the role and functioning of AI will be required.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Government has a central role in ensuring that the benefits of AI are distributed evenly and that the winners compensate the losers created by these technologies. An important aspect of this is the provision of training and alternative employment for individuals and groups who are made redundant by these developments. For this purpose it might be important to use appropriate means of taxation to avoid profits being expatriated while losses remain local.

It will be difficult to regulate AI via straightforward legislation, given the volatile and dynamic nature of this technology. At the same time it will be necessary to continuously refine the definition of what AI should be allowed to do, who should be responsible for the consequences, how these consequences are allocated etc. It therefore seems reasonable to establish an AI regulator that oversees the technology, contributes to the development of standards and best practice and is empowered to enforce such standards. Such a regulator could be similar to the Information Commissioner's Office and would be likely to collaborate closely with the ICO.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

The development of new technologies is not a national matter. The leading tech companies are international players that can easily change jurisdiction. Any intervention by the UK with the aim to render AI beneficial must seek close international cooperation, in the first instance with the EU. The Council of Europe is proposing close cooperation between the Council of Europe, the European Union and UNESCO to develop a harmonised legal framework and regulatory mechanisms at the international level (van Est & Gerritsen, 2017).

The EU has proposed creating a regulator for AI and robotics, the “EU Agency for Robotics and Artificial Intelligence” (Committee on Legal Affairs, 2017). A UK regulator would need to work very closely with this European agency and align and synchronise their principles and activities.

Finally, the UK has a limited capacity to undertake technology foresight and assessment as a service to policymakers. The Parliamentary Office of Science and Technology (POST) provides valuable services, such as the interdisciplinary programme on big data which is relevant to this inquiry. The example of AI and the present call for evidence shows that an independent parliamentary ability to investigate upcoming issues is required in a society that increasingly relies on science and technology to solve the most pressing problems.

Many European countries have more developed technology assessment institutions, such as the TAB in Germany, the Rathenau Instituut in the Netherlands or the Danish Board of Technology. These technology assessment institutions provide advice to national parliaments and are represented in the European Parliamentary Technology Assessment network (<http://www.eptanetwork.org/>). The UK should ensure continued presence in this network post Brexit and should consider strengthening Parliament’s technology assessment capacity.

References

Committee on Legal Affairs, 2017. REPORT with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) (No. A8-0005/2017). European Parliament.

Van Est, R., Gerritsen, J., 2017. Human rights in the robot age - Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality (Report to the Parliamentary Assembly of the Council of Europe (PACE)). Rathenau Instituut, The Hague.

6 September 2017

Ordnance Survey – Written evidence (AIC0090)

The pace of technological change

- 1 What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?
 - 1.1 We define artificial intelligence (AI) as the ability for computer systems to learn from and adapt to new circumstances. AI enables a degree of autonomy, as opposed to automation; in the latter, the parameters and circumstances of operation are known and controlled.
 - 1.2 AI for specific tasks is advancing rapidly. An example of this is autonomous driving, where the pace of progress is very rapid, and we can expect substantial advances towards full autonomy over the next 5-10 years. However, more general AI (that is, a 'thinking machine') is still a remote prospect and is unlikely to be realised within the next 20 years.
 - 1.3 The pace of AI development is affected by the availability of processing power (hardware), algorithms (software) and data, all of which need to be aligned in pursuit of a defined goal. The availability of data – and particularly labelled or tagged data – is often a key limiting factor for the current generation of AI.
 - 1.4 Ordnance Survey (OS) is experiencing the reverse situation. We have an abundance of labelled data – in this instance, rich data describing land parcels and geospatial features such as buildings, roads and railways – but processing power is currently a major challenge at the required scale to robustly develop AI for the classification of raw geospatial data. We are working with a range of organisations to identify how this data can be exploited to yield value in a range of contexts.
 - 1.5 Openness is a critical determinant of AI development; collaboration is generally faster when resources are available as a platform for collaboration and exploitation. Many broad AI algorithms are openly available online¹⁰¹³.
- 2 Is the current level of excitement which surrounds artificial intelligence warranted?

No response.

¹⁰¹³ For example, see <https://github.com/tensorflow/models/tree/master/inception>

Impact on society

- 3 How can the general public best be prepared for more widespread use of artificial intelligence?

Please refer to our response to question 8.

- 4 Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

No response.

Public perception

- 5 Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

- 5.1 We recommend that public understanding should be improved, given the significant likely impact of AI on society. Please refer to our response to question 8.

Industry

- 6 What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

- 6.1 We envisage that the proliferation of data will create the potential for beneficial impacts across all sectors.

- 6.2 For public services, we anticipate that AI will enable a significant shift from reactive to proactive intervention through accurate prediction against real-time information. This capability will also have a significant benefit for risk-based industries such as insurance, but it is likely to also more widely affect risk assessment and mitigation activities of all industries across all sectors. In the context of smarter cities, this capability is likely to offer particular benefits across transport, air quality and public health.

- 6.3 We also envisage that AI realised through connected and autonomous vehicles (CAVs) will have a profound effect on all aspects of mobility and freight, with improved convenience and reduced costs spread throughout all of industry and society; however, this development will have a disruptive impact on the transport industry and supporting national infrastructure as it stands today.

- 7 How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How

can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- 7.1 Data is the 'oil' of AI. For AI processes and outcomes to be effective, unbiased and non-discriminatory, source data needs to have clear provenance and well-described quality.
- 7.2 There is a role for public bodies to own or regulate aspects of digital infrastructure to ensure that AI does not become dominated by a few large corporations. Moreover, digital infrastructure, like its physical counterparts, cannot be relied upon to be developed by market forces alone. Open standards are particularly important in enabling interoperability between digital infrastructure and avoiding lock-in to proprietary technologies.
- 7.3 When data can be exchanged and used as the basis for decision making with confidence and without friction, an authoritative data record about a real-world object – often termed its 'digital twin' – may become valuable in its own right. Evidence for this lies in the rise of (a) Building Information Modelling (BIM), a concept at the foundation of the government's Digital Built Britain initiative¹⁰¹⁴ and (b) the growth of the 'PropTech' industry¹⁰¹⁵. However, creating and maintaining digital infrastructure requires sustained effort and investment. An example of this for critical national infrastructure is the ITRC Mistral¹⁰¹⁶ project, which is enabling strategic analysis and is an example of the kind of data resource that will become increasingly essential for the functioning of AI systems.
- 7.4 AI is becoming a growing feature of online platforms and social media. To date, there is little evidence of public disquiet over the trade-off between enhanced services and reduced privacy. We suggest that government should carefully consider how to preserve the public good and public trust by enabling individuals to assess their privacy status and understand what personal information they have knowingly or unknowingly traded. This is an area where Europe has taken a proactive role, and the UK's exit from the EU merits the UK government's particular focus in this area.
- 7.5 We build on some of these points in our response to question 10.

¹⁰¹⁴ <http://digital-built-britain.com/>

¹⁰¹⁵ PropTech innovation is being fostered by an Ordnance Survey/Land Registry partnership – see <https://geovation.uk/land-registry-partnership/>

¹⁰¹⁶ <http://www.itrc.org.uk/>

Ethics

- 8 What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?
- 8.1 Clearly, the use of AI to guide decision making in any form is associated with many ethical issues. We suggest that to address these, an equivalent to PETRAS¹⁰¹⁷ could be highly valuable. PETRAS is a 3-year research programme focusing on matters relating to privacy, ethics, trust, reliability, acceptability, and security in respect of the Internet of Things (IoT). This programme has been designed to mirror an IoT city demonstrator project, CityVerve¹⁰¹⁸, which is funded by DCMS through Innovate UK. We believe that a similar AI ethics and trust research programme alongside an AI demonstrator project, possibly focussed on government administration, and a public communication initiative would make a valuable contribution in engaging the public and informing the policy landscape in this area.
- 8.2 Location is a fundamental component of personal identity and behaviour. Growing AI capability will make it increasingly feasible to identify an individual from anonymised data by scraping and processing location-based information from various sources. OS is supporting a project, GEOSEC¹⁰¹⁹, which is considering personal location privacy considerations within PETRAS. Personal location data needs to be carefully managed within future AI applications.
- 9 In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?
- 9.1 We suggest that the use of black box AI systems should be judged as more acceptable when they are moderated by a human-based quality assurance sampling process. OS operates the principle of selecting algorithms for image classification that are as open as possible and ensuring that there is a human check involved in the process.

The role of the Government

- 10 What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

¹⁰¹⁷ <https://www.petrashub.org/>

¹⁰¹⁸ <http://www.cityverve.org.uk/>

¹⁰¹⁹ <https://www.petrashub.org/portfolio-item/lightweight-security-and-privacy-for-geographic-personal-data-and-location-based-services/>

- 10.1 Building on our response to questions 7 and 8, we advocate a strong public role in establishing authoritative reference data frameworks to underpin an open, innovative and interoperable digital economy, including AI. The exchange and repurposing of data is central to AI, but for this to be effective, sustainable and socially acceptable the government needs to take a lead in ensuring that reference data frameworks are discoverable and can be described in terms of quality and confidence. This builds on the government’s Industrial Strategy role.
- 10.2 For example: as part of its public task, Ordnance Survey maintains a geospatial data framework which provides a critical infrastructure role in many public and commercial functions. Geospatial represents a ‘golden thread’ that links multiple datasets. With the great majority of all data records having some connection to ‘place’, a geospatial framework built on clear provenance and open standards is vital to integrate and exchange disparate types of information effectively. This makes location and geospatial highly relevant to the future of AI.
- 10.3 A specific example of an effective data framework is the National Address Gazetteer (NAG), which is maintained by GeoPlace, a partnership between Ordnance Survey and the Local Government Association. The NAG is made usable through a variety of addressing products and services which are built on the BS7666 standard and provide an unambiguous and link to the authoritative address via the Unique Property Reference Number¹⁰²⁰. While the address data itself may provide a useful input into an AI for analysis, the UPRN as an authoritative reference provides a means to establish links and correlations between variables coexisting at the same location. This is an example of a publicly-defined and owned data resource which offers benefits across many sectors.
- 10.4 We are exploring the next generation of the frameworks required to support smart city applications through the Manchester CityVerve IoT project and we are also working with the Centre for Connected and Autonomous Vehicles (CCAV) in respect of data exchange requirements relating to CAVs; a consideration which emerged as a key output of the Atlas¹⁰²¹ project.
- 10.5 We consider CCAV to be a useful example of how government has created a specialist cross-departmental unit (in this case, involving BEIS and DfT) to coordinate both the research agenda (in conjunction with funding

¹⁰²⁰ <https://www.geoplace.co.uk/addresses/uprn>

¹⁰²¹ Atlas is a CAV feasibility study funded by Innovate UK: see <https://www.ordnancesurvey.co.uk/business-and-government/smart/mobility-cav.html>.

authorities and research councils) and also the policy response. We suggest that this is a good model to follow for AI.

Learning from others

- 11 What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

No response.

6 September 2017

Marion Oswald and Sheena Urwin – Written evidence (AIC0068)

House of Lords Select Committee on Artificial Intelligence Call for Evidence

Written evidence submitted by:

Marion Oswald (corresponding author)
Senior Fellow in Law and Director of the Centre for Information Rights
University of Winchester

and

Sheena Urwin
Head of Criminal Justice
Durham Constabulary

Introduction

1. This submission results from collaboration between the authors to reflect upon the recent operational deployment of an algorithmic risk assessment tool within Durham Constabulary. The submission also includes aspects of the corresponding author's research (together with Jamie Grace, Sheffield Hallam University) into safeguards required for the responsible use of algorithmic tools within policing, including a freedom of information based study. We have focused upon the questions asked by the Committee which might best be addressed – in full or in part – by the work done during the collaboration to date.

2. In this submission, we define an algorithm as a mathematical formula implemented by technology: 'a sequence of instructions that are carried out to transform the input to the output.' (Al Paydin, 2016) We are concerned with *machine learning* whereby the computer learns and extracts the algorithm for the task from the given input data. (We do not comment on coded rules, programmed logic or database interrogation or linking.)

The pace of technological change: question 1

What is the current stage of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

3. Our answer to this question focuses on the use of algorithmic decision-making tools within the policing and criminal justice context. In the UK policing context, the use of algorithmic decision-making tools could be described as being in a developmental stage, with decisions on implementation being taken on a force by force basis. This contrasts with the United States where algorithmic tools are now used in a number of States across the criminal justice system to inform

human decision-making with respect to decisions or judgements about individuals.

4. It has been suggested (Oswald, Grace, 2016) that there are currently three main purposes for algorithmic data or intelligence analysis within the policing context: i) predictive policing on a macro level incorporating strategic planning, prioritisation and forecasting; ii) operational intelligence linking and evaluation which may include, for instance, crime reduction activities; and iii) tactical decision-making or risk-assessments relating to individuals.

5. One UK force has been reported to be making substantive use of a predictive policing tool developed by the private sector ('PredPol', implemented by Kent Constabulary) in order to predict areas where offences are likely to take place. It has recently been reported that West Midlands police are testing a third party system called 'Valcri' for use in the investigative process, a tool that aims to group similar crimes by the analysis of semantic features. A 2016 freedom of information-based study suggested a relatively small number of UK police forces (14%) were using computational or algorithmic data analysis or decision-making in relation to the analysis of *intelligence*, with tools stated to be used for all three of main purposes mentioned above. (Oswald, Grace, 2016)

6. Durham Constabulary has implemented an algorithmic risk-assessment tool in category iii) above (tactical decision-making or risk-assessments relating to individuals) known as the 'Harm Assessment Risk Tool' (or 'HART'). The tool was initially developed by statistical experts based at the University of Cambridge in collaboration with the force. It has been implemented to *aid* decision-making by custody sergeants when assessing the risk of future offending and so whether an offender is eligible to participate in the force's 'Checkpoint' programme, thus avoiding court (an out-of-court disposal aimed at reducing future offending). In order to divert these offenders away from prosecution, Checkpoint must first identify those who present an appropriate risk of reoffending. The current HART model separates offenders into three different predicted risk groups, only one of which is eligible for the Checkpoint treatment (Moderate risk offenders).

7. It is understood that other UK forces are considering the development of similar tools, although this may be in connection with different programmes or contexts, with potential for such tools to be implemented to prioritise investigative actions or where the police have to decide whether to supply public protection risk information, based on an actuarial judgement (such as 'Clare's Law').

8. Deployments by police forces of Big Data and algorithmic technologies may be hindered by the localised structure of UK policing and the lack of compatibility between, and fragmented ownership of, key police databases (Babuta, 2017). There are, equally importantly, significant legal and ethical issues to be addressed. We comment upon some of these below.

Impact upon society: question 3

How can the general public best be prepared for more widespread use of artificial intelligence?

and

Public perception: question 5

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

9. From the beginning of the development of this tool, and since its validation, Durham Constabulary has been open about its use of HART, the tool's internal workings and the results of the first validation exercise, attracting considerable attention (and sometimes criticism). The purpose of being so open was to acknowledge that this approach is new to policing and is therefore also new to communities. Secondly, being open permits learning and understanding from others in relation to concerns and issues that exist. Thirdly and lastly, capturing that learning throughout the exploratory process has allowed the Constabulary to use these lessons to help develop a framework to support and assist other police organisations – 'Algo-care' (we expand further upon this in answer to the Ethics questions below).

10. We believe that public awareness, understanding and thus preparedness for the more widespread use of artificial intelligence within policing can only be helped by such an open approach. We appreciate though that such public openness may not be possible with all tools deployed in the investigatory context, where openness about techniques may be damaging to the investigation of crime or national security. In addition, the details of the HART model are complex (for instance, it contains over 4.2 million decision points, all of which are highly interdependent on the ones that precede them within the 'tree' structure). These details could be made available; however in doing so we risk providing complex information and an illusion of transparency being nurtured - referred to as a 'transparency fallacy' (Edwards and Veale, 2017). In an effort to be open the response is an open approach, to the extent that the information is not meaningful in terms of a right to explanation. Comprehensive oversight by an independent, expert body is likely also to be needed to provide the appropriate reassurance and to avoid costly and protracted challenges to systems' use in court.

11. As we have commented upon elsewhere (Oswald, Grace, Urwin & Barnes, 2017), one of the issues regarding the use of algorithms by the police may be the lack, as yet, of definitive answers to the question of benefits and harms. The deployment of algorithmic technology may be in many ways, experimental. These nuances and uncertainties need to be better understood. This could be achieved by requiring either i) reports and communications from an independent oversight body as referred to above; and/or ii) the inclusion of members of the public and representatives of the third sector or campaigning organisations in membership of ethical review bodies, such as the Cleveland & Durham Joint

External Ethics Committee or the National Statistician's Data Ethics Advisory Committee.

12. In addition, consideration could be given to expanding the sector specific model publication schemes for public authorities pursuant to the Freedom of Information Act 2000 s20 to include appropriate information about the use of algorithmic decision-making tools in order to encourage such information to be provided publicly and proactively (and thus to set expectations for third party providers).

Ethics – question 8

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

13. We have recently commented in depth upon the legal, societal and ethical issues raised by the use of algorithms in policing, using HART as an illustration (Oswald, Grace, Urwin & Barnes, 2017). In summary, HART raises: i) issues of *uncertainty* (its outputs are probable but not conclusive and it cannot assess all possible relevant factors); ii) issues of *opacity*, particularly important when a tool is being used to support a decision-making process in the criminal justice context; iii) issues of possible *bias* in relation to the use of historical offender data, and residential postcode, as inputs; iv) questions of *value-judgements* built into the operation of the algorithm, in the case of HART relating to the 'trade-off' to be made between false positives and false negatives in order to avoid errors that are thought to be the most dangerous: in this context, offenders who are predicted to be relatively safe, but then go on to commit a serious violent offence (high risk false negatives); v) the question of the *long term effect* of algorithmic tools on human decision-making processes within a police force.

14. The above issues link closely to matters of law and in particular those related to judicial review and human rights principles. Care must be taken to ensure that an algorithmic tool is not taking into account irrelevant factors, or is not, in practice, becoming the decision-maker rather than decision-support, thus fettering the public body's discretion. A substantial risk exists that humans become afraid to challenge computer-aided decision making. In addition, rules of natural justice and the duty to give reasons mean that public bodies with significant power over the lives of individuals must take steps to foster meaningful transparency, in ways that would allow a defendant to challenge the operation of the tool, and for a solicitor to provide meaningful advice to his client about how to approach questioning. From a human rights perspective, the use of an algorithm must be 'necessary' and 'proportionate', and its use in determining an individual's liberty must be foreseeable. However, the experimental nature of developing technology may cause difficulties in the assessment of the proportionality of a particular algorithm, as it may be too early to assess the benefits and harms conclusively.

15. **Therefore, we suggest two linked concepts:** i) a concept of 'experimental' proportionality and ii) a decision-making guidance framework for

the deployment of algorithmic assessment tools in the policing context called 'ALGO-CARE'. 'Experimental' proportionality would have the dual advantages of permitting the use of unproven algorithms in the public sector in order that benefits and harms can be fully explored, yet giving the public confidence that such use would be independently controlled and time-limited and the proportionality subject to a further review on a stipulated future date (so a similar aim to a 'sunset' clause in legislation).

16. This concept would encapsulate two elements. First, a formal approach giving the 'benefit of the doubt' to the public sector body where it is not yet possible to determine with any certainty the balance or imbalance of benefits and disadvantages in relation to the new algorithmic technology. Secondly, a change to statutory procedure and forms of relief available so that the High Court could order that the benefits and harm risks, and hence the proportionality of the particular use of the algorithm, be reviewed in another hearing after a period of time (an approach that could also be taken by a regulator or oversight body). The police force, or other public sector body, would still be required to comply with the requirements of natural justice, even in the 'experimental' stage. In addition, the public sector body must still demonstrate a baseline connection to a legitimate aim and that the outcomes and benefits (even if these are as yet theoretical or only foreseen) are rationally connected to that aim and, based on the knowledge available, a reasonable belief that there is not an excessive cost to human rights. The role of a suitable senior officer, aware of the algorithms detail, to interpret individual results and ensure that contextual factors are considered cannot be underplayed in this proposed experimental stage.

17. Our second linked proposal is designed to promote decision-making consistency and rigour: a decision-making framework called '*Algorithms in Policing –Take ALGO-CARE™*'. The framework reflects the experience of Durham Constabulary in developing and rolling out its algorithm associated with the Checkpoint programme. It also aims to translate key public law and human rights principles into practical considerations and guidance that can be addressed by public sector bodies.

18. While the authors note that a number of organisations are developing, or advocate developing, other high level principles in respect of algorithms and A.I. (which can be helpful to represent ethical norms and in setting a general direction of travel), we would submit that they often do not provide enough practical certainty for the development of administrative and assessment frameworks (or for practitioners to refer to in their day-to-day work). *Algo-care* aims to address these concerns, and to provide a decision-making framework that could work in different policing contexts, and potentially more widely across the public sector.

19. The current working version of '*Algorithms in Policing –Take ALGO-CARE™*' is set out in the Appendix, together with additional explanatory notes. Each word in the mnemonic – *Advisory; Lawful; Granularity; Ownership; Challengeable;*

Accuracy; Responsible; Explainable – is supplemented by questions and considerations representing key legal considerations (such as necessity and proportionality, and natural justice and procedural fairness), as well as practical concerns such as intellectual property ownership and the availability of an ‘expert witness’ to the tool’s functionality. It is intended for use by senior practitioners and decision makers as well as those developing algorithms at a working level.

Ethics – question 9

In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable? When should it not be permissible?

20. We would suggest that concerns around transparency and accountability cannot be addressed in a one-size-fits-all way. It would not be appropriate, for instance, for the functionality of a tool used in the investigative process to be ‘transparent’ in the sense of the detailed functionality being publicly available. Where a tool assists in the decision-making about an out-of-court disposal, however, information about its use should be made available to the affected individual and/or his legal adviser (to address ECHR Article 6 concerns).

21. *Algo-care* identifies that, at the very least, the public body should be able to explain the decision-making rule(s) and the impact that each factor has on the final score or outcome, and ensure that it has access to and can deploy a data science expert to explain the algorithmic tool (in a similar way to an expert forensic pathologist). The framework also notes that development specifications should incorporate as appropriate the latest methods of interpretable, interactive and accountable machine learning systems (see for instance Kroll et al., 2017).

22. These factors necessitate the careful drafting of procurement contracts with third party software suppliers (commercial or academic). Contracts should require disclosure of the algorithmic workings in a way that would facilitate investigation by a third party in an adversarial context if necessary (and the provision of an expert witness/evidence of the tool’s operation). It is our view that commercial confidentiality should not be permitted to be a barrier to appropriate scrutiny.

23. In addition to appropriate rights to use, amend and disclose the software tool, public sector bodies should pay attention to rights over any third party data that have been used as inputs, such as mosaic postcodes. Although *Algo-care* identifies that open source software as default should be considered, it is appreciated that access to the source code does not necessarily, of itself, result in an appropriately understandable and challengeable tool. Such access could however aid validation exercises for accuracy and bias.

The role of the Government – question 10

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

24. The Government can take a crucial role in ensuring clarity in relation to the legal framework, oversight mechanisms and appropriate sectorial guidance governing the development and use of artificial intelligence by the public sector. Our *Algo-care* framework recommends that ethical considerations, such as consideration of the public good and moral principles are also factored into the deployment decision-making process. The Government could additionally play a role in ensuring that administrative arrangements such as ethical review committees incorporating independent members are established for such a purpose.

25. We would advocate that the Government should take a role in identifying the categories of decision – such as those that may impact Article 2 rights or the fundamentals of a fair trial – that would not benefit from ‘experimental’ proportionality and indeed which should be excluded from the purview of artificial intelligence altogether, at least for the time being.

26. For further details, please refer to Oswald, Grace, Urwin and Barnes (2017) ‘Algorithmic risk assessment policing models: Lessons from the Durham HART model and ‘Experimental’ proportionality’ available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029345

Marion Oswald and Sheena Urwin
4 September 2017

References

Ethem Alpaydin, ‘Machine Learning’ (MIT Press, 2016)

Alexander Babuta ‘Big Data and Policing: An Assessment of Law Enforcement Requirements, Expectations and Priorities’ Royal United Services Institute Occasional Paper, 6 September 2017

Lilian Edwards and Michael Veale ‘Slave to the Algorithm? Why a ‘Right to Explanation’ is Probably Not the Remedy You are Looking for’ (May 23, 2017). Available at SSRN: <https://ssrn.com/abstract=2972855>

Kroll et al., ‘Accountable Algorithms’ 165 U Pa. L. Rev. 633 (2017)
<https://www.pennlawreview.com/print/?id=553>

Oliver Moody, ‘Detectives call in AI to hunt offenders’ The Times, May 17, 2017
<https://www.thetimes.co.uk/edition/news/detectives-call-in-ai-to-hunt-offenders-8g5ncqxsr>

Marion Oswald, Sheena Urwin, Jamie Grace and Geoffrey Barnes ‘Algorithmic Risk Assessment Policing Models: Lessons from the Durham Hart Model and ‘Experimental’ Proportionality’ (August 31, 2017). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029345

Marion Oswald and Jamie Grace, ‘Intelligence, policing and the use of algorithmic analysis: a freedom of information-based study’ (2016) Vol 1, No. 1, Journal of

Information Rights, Policy & Practice
<https://journals.winchesteruniversitypress.org/index.php/jirpp/article/view/16>
<http://www.predpol.com/how-predpol-works/>

Appendix - Algorithms in Policing –Take ALGO-CARE™

<ul style="list-style-type: none"> Algorithms in Policing – Take ALGO-CARE™ A proposed decision-making framework for the deployment of algorithmic assessment tools in the policing context 		
<ul style="list-style-type: none"> A 	<ul style="list-style-type: none"> Advisory 	<ul style="list-style-type: none"> Is the assessment made by the algorithm used in an advisory capacity? Does a human officer retain decision-making discretion? What other decision-making by human officers will add objectivity to the decisions (partly) based on the algorithm?
<ul style="list-style-type: none"> L 	<ul style="list-style-type: none"> Lawful 	<ul style="list-style-type: none"> On a case-by-case basis, what is the policing purpose justifying the use of algorithm, both its means and ends? Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes? In what way will the tool improve the current system and is this demonstrable? Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim? Is the operation of the tool compliant with national guidance?
<ul style="list-style-type: none"> G 	<ul style="list-style-type: none"> Granularity 	<ul style="list-style-type: none"> Does the algorithm make suggestions at a sufficient level of detail/granularity, given the purpose of the algorithm and the nature of the data processed? Is data categorised to avoid 'broad-brush' grouping and results, and therefore issues potential bias? Do the benefits outweigh any technological or data quality uncertainties or gaps? Is the provenance and quality of the data

		<p>sufficiently sound? Consider how often the data should be refreshed. If the tool takes a precautionary approach towards false negatives, consider the justifications for this.</p>
<ul style="list-style-type: none"> • O 	<ul style="list-style-type: none"> • Ownership 	<ul style="list-style-type: none"> • Who owns the algorithm and the data analysed? Does the force need rights to access, use and amend the source code and data analysed? How will the tool be maintained and updated? Are there any contractual or other restrictions which might limit accountability or evaluation? How is the operation of the algorithm kept secure?
<ul style="list-style-type: none"> • C 	<ul style="list-style-type: none"> • Challengeable 	<ul style="list-style-type: none"> • What are the post-implementation oversight and audit mechanisms e.g. to identify any bias? Where an algorithmic tool informs criminal justice disposals, how are individuals notified of its use (as appropriate in the context of the tool’s operation and purpose)?
<ul style="list-style-type: none"> • A 	<ul style="list-style-type: none"> • Accuracy 	<ul style="list-style-type: none"> • Does the specification match the policing aim and decision policy? Can the stated accuracy of the algorithm be validated reasonably periodically? Can the percentage of false positives/negatives be justified? How was this method chosen as opposed to other available methods? What are the consequences of inaccurate forecasts? Does this represent an acceptable risk (in terms of both likelihood and impact)? Is the algorithmic tool deployed by those with appropriate expertise?
<ul style="list-style-type: none"> • R 	<ul style="list-style-type: none"> • Responsible 	<ul style="list-style-type: none"> • Would the operation of the algorithm be considered fair? Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review alongside other IT developments in policing?

		<ul style="list-style-type: none"> • Would it be considered to be for the public interest and ethical?
<ul style="list-style-type: none"> • E 	<ul style="list-style-type: none"> • Explainable 	<ul style="list-style-type: none"> • Is appropriate information available about the decision-making rule(s) and the impact that each factor has on the final score or outcome (in a similar way to a gravity matrix)? Is the force able to access and deploy a data science expert to explain and justify the algorithmic tool (in a similar way to an expert forensic pathologist)?

<ul style="list-style-type: none"> • <i>Brief explanatory notes and additional considerations</i> • <i>The Algorithms in Policing – Take ALGO-CARE™ framework is intended to provide guidance for the use of risk-assessment, predictive, forecasting, classification, decision-making and assistive policing tools which incorporate algorithmic machine learning methods and which may impact individuals on a micro or macro level</i> 		
<ul style="list-style-type: none"> • A 	<ul style="list-style-type: none"> • Advisory 	<ul style="list-style-type: none"> • Care should be taken to ensure that an algorithm is not inappropriately fettering an officer’s discretion, as natural justice and procedural fairness claims may well arise. Consider if supposedly advisory algorithmic assessments are in practice having undue influence. If it is proposed that an algorithmic decision be automated and determinative, is this justified by the factors below? Data protection rights in regard to automated decisions may then apply.
<ul style="list-style-type: none"> • L 	<ul style="list-style-type: none"> • Lawful 	<ul style="list-style-type: none"> • The algorithm’s proposed functions, application, individual effect and use of datasets (police-held data and third party data) should be considered against necessity, proportionality and data minimisation principles, in order to inform a ‘go/no-go’ decision. In relation to tools that may inform criminal justice disposals, regard

		should be given to the duty to give reasons.
• G	• Granularity	<ul style="list-style-type: none"> • Consideration should be given to common problems in data analysis, such as those relating to the meaning of data, compatibility of data from disparate sources, missing data and inferencing. Do forces know how much averaging or blurring has already been applied to inputs (e.g. postcode area averages)?
• O	• Ownership	<ul style="list-style-type: none"> • Consider intellectual property ownership, maintenance of the tool and whether open source algorithms should be the default. When drafting procurement contracts with third party software suppliers (commercial or academic), require disclosure of the algorithmic workings in a way that would facilitate investigation by a third party in an adversarial context if necessary. Ensure the force has appropriate rights to use, amend and disclose the tool and any third party data. Require the supplier to provide an 'expert' witness/evidence of the tool's operation if required by the force.
• C	• Challengeable	<ul style="list-style-type: none"> • The results of the analysis should be applied in the context of appropriate professional codes and regulations. Consider whether the application of the algorithm requires information to be given to the individual and/or legal advisor. Regular validation and recalibration of the system should be based on publicly observable (unless non-disclosable for policing/national security reasons) scoring rules.
• A	• Accuracy	<ul style="list-style-type: none"> • How are results checked for accuracy, and how is historic accuracy fed back into the algorithm for the future? Can

		forces understand how inaccurate or out-of-date input data affects the result?
<ul style="list-style-type: none"> • R 	<ul style="list-style-type: none"> • Responsible 	<ul style="list-style-type: none"> • It is recommended that ethical considerations, such as consideration of the public good and moral principles (so spanning wider concerns than legal compliance) are factored into the deployment decision-making process. Administrative arrangements such as an ethical review committee incorporating independent members could be established for such a purpose (such as Cleveland & Durham Joint External Ethics Committee¹⁰²² or the National Statistician’s Data Ethics Advisory Committee).¹⁰²³
<ul style="list-style-type: none"> • E 	<ul style="list-style-type: none"> • Explainable 	<ul style="list-style-type: none"> • The latest methods of interpretable and accountable machine learning systems should be considered and incorporated into the specification as appropriate. This is particularly important if considering deployment of ‘black box’ algorithms, where inputs and outputs are viewable but internal workings are opaque (the rule emerges from the data analysis undertaken). Has the relevant Policing & Crime Commissioner been briefed appropriately?

© Marion Oswald, Jamie Grace

4 September 2017

¹⁰²² <https://www.durham.police.uk/About-Us/Transparency-and-Integrity-Programme/Pages/1-Oversight-and-Accountability.aspx>

¹⁰²³ <https://www.statisticsauthority.gov.uk/national-statistician/national-statisticians-data-ethics-advisory-committee/>

Professor Maja Pantic – Written evidence (AIC0215)

Maja Pantic

Imperial College London, Computing Department

Artificial Intelligence (AI) is a set of computational techniques used to represent knowledge in formats usable by computers (machines) and to solve problems that require intelligence if solved by people. This set of techniques is large and versatile and includes pattern recognition, computer vision, audio processing, knowledge and data representation methodologies, data mining, machine learning, robot sensing and learning, planning and reasoning methodologies, dialogue management techniques, etc. The application areas of AI span and impact all sectors, economies, and people. AI is not only changing what and how we are doing business, but also who we are (think of the impact of social media and dating apps on how we connect with friends, family, and future partners).

(1) The pace of technological change: Until recently, the use of AI and intelligent digital devices (e.g. robots) was confined to tightly controlled tasks in laboratory environments and very few specific industries. Nowadays, ubiquitous online connectivity, dramatic increase in computational power, availability of data, and advances in sensors and machine learning algorithms enable deployment of AI technologies across all sectors and for a wide range of tasks including autonomous vehicles, drones, online advisors, smart home appliances, robotic toys for autistic kids, automatic surveillance, etc. AI technology is penetrating all industrial sectors and people use it all the time (think of the super computer in your pocket called mobile phone). At the same time, AI is becoming more adaptive and flexible, and more bio-inspired. Therefore, it is likely that the next generation of AI technologies will be increasingly focused on tightly-coupled human-machine collaboration. This will bring great benefits such as helping doctors to make medical diagnoses based on intelligent and more accurate sensors and using AI-empowered search engines that can retrieve relevant cases from all global medical databases, or having smart homes and home robotic companions that would enable independent living in a very old age, or having intelligent drones and robots for all hazardous situations (fire, earthquake, floods, etc.), to mention but a few. But, in the next 5-7 years, this will also usher major economic, social, and cultural changes (see point 2). It also raises many ethical and psychological questions (see point 4).

(2) Impact on society: The scale and breadth of economic, social and cultural changes that the AI developments will bring about are of such phenomenal proportions that they cannot be fully envisaged.

Economy: AI introduces disruptions that are unexpected, profound, and happen incredibly fast. The ubiquitous iPhone was introduced in 2007. Yet there will be over 2 billion iPhones sold by the end of 2017. In 2010 Google Driverless Car

program was launched and in 2015 Tesla's Autopilot feature was introduced. Already in May 2017 Wired magazine listed 263 companies working on driverless cars. Disruptors like Airbnb, Uber, and Amazon, all based on very simple AI principles, were relatively unknown just a few years ago. Nowadays, Airbnb is worth 30% than the largest hotel chain while having only 2370 employees and owning 0 bedrooms. These staggering returns-to-scale ratios are consistent across all AI-oriented companies. For example, the world's largest company in terms of return is Walmart, a traditional general merchandiser with a \$214k return per employee. Amazon is an AI-based general merchandiser and has \$433k return per employee. The fact that a unit of wealth is created by AI-based businesses with much fewer workers is possible because AI-based businesses have marginal costs that tend towards zero (in case no physical merchandise is involved, e.g., as in Airbnb, Instagram, Facebook, etc.). This makes AI-based business very attractive. Furthermore, because smart sensors are being installed everywhere including our streets, houses, cars, cloths, and pets, and because of ubiquitous online connectivity of people and businesses, **AI-based disruptions can be expected in any branch of industry and service provision** (including banking, real estate, news broadcasting and health care). However, many companies and governments do not seem prepared for this.

Employment & Skills: So far, the evidence is this: AI-based businesses employ much fewer people than the same businesses run in a traditional way (think of the above-mentioned examples like Airbnb and Amazon). A further evidence is that as AI continues to develop, more and more jobs are being automated. Telemarketers, cashiers, librarians, post-office workers and couriers, are already fully automated jobs, rarely performed by a human. Long-distance drivers, tax preparers, car damage appraisers, farm workers, and real estate brokers are some of the jobs that are highly likely to become fully automated in the next 5 years. Jobs that rely on intrinsically human traits like empathy (psychologists, nursery workers, geriatricians, etc.) cannot be replicated by machines and will be in high demand due to the growing elderly population. New jobs that will be created in the coming years will all require complex problem solving skills and knowledge of AI and digitalisation. In turn, many people will need to be reskilled. Also, most of the future jobs will be of an on-demand type, where workers are contractors, working online from home (or from another country). The downside is that workers no longer enjoy job security. On the other hand, aided by decentralised digital payment systems (like bitcoin), this introduces unforeseen **challenges to tax collection**.

Innovation capital, Concentration of the innovation capital, and Brian drain: Given that the future world would be largely digitalized and based on AI, the great beneficiaries would be AI innovators and providers of relevant intellectual and innovation capital. Academia was always regarded as the foremost place to engage with innovation. Academia and educational institutions were also regarded as the foremost places to learn new skills. However, salaries, carrier incentives, and research funding schemes provided by the government to the

academia, are incomparable less attractive than those offered by the AI industries, especially the AI giants like Google, Facebook, Apple, and Amazon. Consequently, schooled AI researchers – graduated MSc students, PhD students, and Professors – leave academia in great numbers to work for one of those AI giants. This **brain drain** from academia results in two major drawbacks: (1) academia is left with no new generation of AI researchers who could continue AI research in public domain and actively contribute to schooling of our children and reskilling of our people, and (2) the four AI industrial giants listed above amassed intellectual and innovation capital, with the fair prospect of owning 90% of all innovation in the years to come. This introduces numerous risks including services and product market monopoly by the AI industrial giants, education market monopoly by them, and power balance disruption (power shift from the state and the government to a few global companies), to mention but a few.

Government: Ultimately, it is the ability of governments to adapt to the AI-based age that is coming that will determine their survival and the prosperity prospects of their people. Governments must adapt to the fact that power is shifting from state to non-state actors. On one side, AI industrial giants are introducing market monopoly (as explained above) and increasing inequality between those who are technically apt and those who are not (due to a huge salary disbalance introduced by these AI industrial giants). This may cause great social unrest. On the other side, AI and digitalisation may enable citizens to voice their opinions in new ways, coordinate themselves, and possibly circumvent government supervision. Hence, centralised, oppressive, manipulative, and rigid governance would have great difficulties to survive (double-checking of all statements for falsehood will be instant; this AI research topic attracts large interest and investment for already 2-3 years). I believe that governments will be increasingly seen as public-service centres that are evaluated by their ability to deliver promised services in the most efficient and personalised way. The governments will need to come up with legislation to decrease the pay gap that the AI industry giants introduced, to protect from market monopoly that the AI industry giants are increasingly introducing, and to contra-act digital exclusion and digital divide (all people need to be able to be a part of digital economy, AI-based world, and e-governance), to mention but a few issues that the governments will need to address. Governing the **diffusion of innovation** (rather than allowing its concentration into just few AI industrial giants) is the key to mitigate dramatic disruptions and negative effects that AI may bring about. However, it seems that large majority of current governments go about AI as it is business as usual, having policies on AI that are inadequate or absent altogether.

Increased inequality: Digital divide is a pressing issue. There are still many people that do not have access to Internet and even more people who are technically inapt. This limits the ability of those people to participate in digital economy and AI-based world, use online services (medical and governmental), and be employed for jobs involving the use of digital technologies. This divide is

deepen even further by the AI industry giants who introduced humongous **pay gaps** between those who have knowledge of AI and those who do not. Another exacerbated inequality introduced by the steady movement of the economy and the world towards the AI realm, is the increased gender gap. The UK has the lowest percentage of female computer engineering professionals in Europe, at less than 10%. The cumulative effect of (i) low percentage of women involved in computing and (ii) automation of high percentage of jobs that have traditionally given women access to the labour market (tele-operators, cashiers, librarians, secretaries, etc.) is a critical concern, profoundly **widening the gender gap**¹⁰²⁴.

(3) Public perception: The AI is often portrayed by the media as an almighty villain that will soon wipe human civilization from the face of the Earth either by starving humans (by taking all jobs) or by engaging in a robot-led war. This is aided and abetted by public figures who have little or no knowledge of AI (e.g. Stephen Hawking who is a physicist and cosmologist, Elon Mask who has degrees in physics and economics) and whose imperative is personal publicity which brings them additional investments and fame. AI is not anywhere near the capabilities depicted in popular Sci-Fi movies like Ex-Machina (2015), Her (2013), I Robot (2004), Minority Report (2002), or even Blade Runner (1982). The current state of the art in AI is best represented by two examples: a go player (highly data intensive task but in a very limited and simple space) and a robotic vacuum cleaner (automatic sensing and mapping task to enable operating in a relatively small but not predefined space). Yet a very large majority of people believe that the SciFi movies depict the state of affairs in the AI. Consequently, they believe that robots are or will soon become conscious and capable of self-preservation (and, hence, of war against humans). People do not understand that the research has not a slightest clue of what “consciousness” really means and how it is modeled in the human brain. Nonetheless, because of being misinformed, people have aversion to learn about AI (it’s too complex and it is against humanity anyhow), aversion to use AI (although they use iPhones and fly by planes all the time, not understanding that these are less-complex forms of AI), and fear AI (as robots will come and kill us all). It is therefore critical to invest effort and attention to create positive and hope-filled narratives that will inform people on the true opportunities and challenges that AI brings. There are many positive impacts of AI (e.g. to help disabled people and atypically developing children, to make partial medical diagnoses that are more accurate than human can make, to help in all hazardous situations, etc.). People, industries, and governments, could all be empowered by AI. The history also teaches that the extent to which society embraces technological innovation is a major determinant of progress. It is therefore essential that the government, academia, and private sector, work with media to create a **positive and**

¹⁰²⁴ My views on the problems we face due to an increasing gender gap in the AI era to come can be read in the CityA.M. article that I wrote: <http://www.cityam.com/267032/calling-all-women-londons-thriving-tech-sector-needs-you>

truthful narrative about AI, informing people on how they can participate in and benefit from the changes that the AI brings about.

(4) Ethics & Impact of the Individual:

Data privacy: One of the greatest challenges posed by the perpetual online connectivity and AI-based devices and apps is the data privacy. Our credit card numbers, birth dates, how we search the net, how much we pay online for a product or a service, what we are mostly interested in, etc., is all available online and often stored in so-called “cookies”. This information is further used for targeted marketing and targeted price-formation, which may result in financial losses. A large majority of people in the Western World own a smartphone (AI-based mobile phone) and never part with it. This allows tracking of an individual’s every step (through “free” location services and health-monitoring services) and survey his or her life patterns including exercising, eating (where and with whom), usual routes, etc. These data and all information that people post on Facebook, chats they have via WhatsApp (also owned by Facebook), pictures they post on Instagram, etc., become the property of the companies providing the service. This private data is then (i) used to train machine learning techniques that can recognise people by the face, learn their life patterns, and learn their preferences, and (ii) sold to various companies and institutions for their marketing campaigns or any other (more alarming) purpose. It is therefore critical to invest in a multi-stakeholder collaboration of governmental, academic, and social law experts to create a solution for **digital data protection** (a way to “stamp” each datum with a unique personal stamp and allow its usage only to those with whom there is an agreement for usage).

Autonomous robots and driverless vehicles: Autonomous robots including various driverless vehicles are on the rise. Except of airplanes and trains, some of which are currently fully automated and have a human operator mostly for the emergency cases, it is expected that within 5 to 10 years we shall also have fully autonomous service drones, long-distance trucks, boats, and maybe even cars. Autonomous robotic baristas, robotic waiters, and robotic cooks are also all expected to come to the market in the next 10 years. Perpetual faultless operation cannot be expected from any machine and, hence, it is inevitable that autonomous robots will cause material damage and injure human beings at some point. This raises a number of **ethical and legal questions** including (i) whether to allow driverless vehicles at all in urban areas where there is no infrastructure that can guarantee zero loss of human lives, (ii) how to program the robots to cause least damage (this is especially challenging with driverless vehicles that may cause loss of human and animal lives), and (iii) who will be liable for the damages (owner of the robot, producer of the robot, or an insurance company). An open in-depth debate with various stakeholders across governmental, legal, social, national, industrial and academic boundaries is needed to unfold and address these questions.

Auditable AI: The majority of current AI-based algorithms is based on applying machine learning techniques to the relevant data to learn the patterns occurring in the data. The problem here is a twofold (i) the data may be biased and the AI may then learn to unintentionally discriminate people (e.g. favouring people of certain race, religion, etc.), and (ii) the machine learning methods applied may use a “black-box” approach and produce results that cannot be explained which again can unintentionally discriminate people or be intentionally impregnated with software spyware and viruses that can go undetected (e.g. deep learning currently used by 90% of the AI community is a fully “black-box” approach). There is currently a surge of interest by the AI research community to work on methods capable of producing auditable AI and capable of checking the audibility of AI. These efforts should be strongly supported. As impartiality of AI-based algorithms is a must, and **cyber warfare** is a realistic threat, a national and global framework is needed to **govern AI audits** and mitigate these risks. Again, I feel that initiatives and governmental steps towards this goal are absent altogether.

Wellness in AI-based world: More and more people spend more and more time using their smartphones and being online. They meet friends and dates in cyberspaces (20% of nowadays marriages have been started via an online dating site), they read online news, and 90% of UK teenagers never part with their smartphones, spending more than 4 hours engaged with AI and digital content every day. Most of UK children nowadays spend 3 months full-time watching digital content before they reach the age of one! The direct consequences of this are (i) severe **reduction in attention span**¹⁰²⁵, (ii) **shallower cognitive capabilities** and ceasing to control the focus of attention¹⁰²⁶ (this is also why “mindfulness” courses and therapies became such a hype as they try to revert this consequence of being consumed by digital and AI-world), and (iii) **loss of identity** and tendency to live in non-existing ideal worlds pictured on social media. The latter leads to many illnesses including eating disorders (caused by fashion icons and their outlook), and serious mental illnesses including suicidal depression (the number of suicides by teenagers in UK has drastically risen in the later years¹⁰²⁷). As already mentioned in point (3) above, it is therefore critical to invest effort and attention to create narratives that will inform people on the true opportunities and challenges that AI brings. These narratives need to be positive and hope-filled but they also need to clearly enumerates the addictive nature of AI and dramatic effects it can have on our wellness may we succumb to this addictive side of the AI.

¹⁰²⁵ <http://www.medicaldaily.com/human-attention-span-shortens-8-seconds-due-digital-technology-3-ways-stay-focused-333474>

¹⁰²⁶ http://www.nicholascarr.com/?page_id=16

¹⁰²⁷ <http://www.express.co.uk/news/uk/769066/suicide-children-teen-Prince-William-Prince-Harry-Duchess-Cambridge-Heads-Together>

Professor Maja Pantic – Written evidence (AIC0215)

Please let me know if any further explanation is needed regarding any of my opinions expressed above. I shall be more than happy to provide further explanation and evidence in person or in writing.

Dr Maja Pantic, FIEEE, FIAPR, FBCS
Professor of Affective and Behavioural Computing
Department of Computing
Imperial College London

12 September 2017

Dr Andrew Pardoe – Written evidence (AIC0020)

What are the implications of artificial intelligence?

Introduction

Dr Andy Pardoe has a doctorate in artificial intelligence and is the founder of Informed.AI, a community of websites for AI supporting those interested to learn more or working in the industry. Listed by IBM Watson in the Top 20 of Global AI Influencers, listed just below Elon Musk. He runs the annual global achievement awards for Artificial Intelligence. A member of the British Computer Society Special Group for Artificial Intelligence committee. Currently work for Credit Suisse designing their own Machine Learning Platforms and Applications. Andy is an International Speaker on Artificial Intelligence, Machine Learning and Robotic Automation.

These are my own personal views.

Questions to be address

- The current state of artificial intelligence.
- The pace of technological change and the development of artificial intelligence
- The impact of artificial intelligence on society
- The public perception of artificial intelligence
- The sectors most, and least likely, to benefit from artificial intelligence
- The data-based monopolies of some large corporations
- The ethical implications of artificial intelligence
- The role of the Government and
- The work of other countries or international organisations.

1. The current state of artificial intelligence.

While the field of AI has been around since the 1950s, it is only in the last 5 to 10 years that a number of factors have come together to establish an unstoppable progression towards super intelligence and the singularity. Those factors being, (1) the advent of large quantities of data facilitated by big data technologies like Hadoop and Big Table, (2) via Cloud platforms access to scalable computing infrastructure, (3) via the gaming industry multi-processor technologies in the form of graphical processing units (GPUs) and lastly (4) advanced learning algorithms and topologies that enable deep learning neural networks and reinforcement learning. Currently most applications of AI are what is call Narrow or Specific AI (meaning a very targeted application) which a number of companies and research institutions are working on AGI (artificial general intelligence) which holds the promise of delivering super intelligence.

However, most knowledgeable commentators believe that AGI and Super Intelligence will not happen for another 20-40 years.

2. The pace of technological change and the development of artificial intelligence

As detailed by the World Economic Forum, we are at the start of the fourth industrial revolution, and what is most evident is that the rate of change for technology development is increasing at an exponential rate, and we are at the inflection point, where we will see in the next 10-15 years the same amount of change as we have seen in the last hundred years. It is clear to see that the next 25-50 years will be an amazing time of change and transformation. The level of complexity and intelligence shown by our machines will be astonishing and that of science fiction for many. With so many different technologies (Robotics, 3D printing, crypto-currencies and block chain, Virtual and Augmented Reality) becoming mainstream and being integrated together the possibilities are endless for the enabling technology of AI.

3. The impact of artificial intelligence on society

As with any industrial revolution the impact on society is immense, and will be perceived as both positive and negative, The challenge with the fourth industrial revolution is that its potential to impact every single profession and industry, large and small. Even those industries that currently appear to have a low dependency on technology are also in scope for impact. Both low and high skilled professions will suffer from job displacement, and the need for a universal basic income is real. Highly skilled professionals will learn their trade from AI agents in the future, as the normal method to learn wont be available as AI agents will be performing the basic functions that were previously used to learn the trade. Every role will be augmented to some degree by AI agents. In the near future entire companies could be fully automated, with only a handful of human staff who own the company, and deal with operational changes needed, with all other tasks being managed by computer and robotics. Another aspect impacting society will be the social interactions of intelligent robots and androids with humans, and more specifically how humans will consider and react to andriods.

4. The public perception of artificial intelligence

As detailed by the Royal Society Report on AI that surveyed the public, many do not know the specific details of AI. but most understand the applications of AI, including self-driving cars and personal assistants on mobile phones. Overall the perception of AI has been positive with the general public, but there is a high risk of misunderstanding of its abilities and capabilities due to the complexity of the techniques and the fact that the field is still relatively new and frequently improving. The concerns around how we manage AI ethics, safety and legislation

together with areas like autonomous weapons put at significant risk the positive perception of AI with the public.

5. The sectors most, and least likely, to benefit from artificial intelligence

All sectors will benefit from AI, just some will benefit sooner than others. The sectors that will be slower to benefit from AI will be those that are highly manual and / or creative in nature. Will we want to watch a theatre production of Shakespeare's Hamlet performed by robots. One of the unseen beneficiaries of AI are Charities, as advance data analytics can help a charity better target its services to those in most need, companies like DataKind are facilitating such activities.

6. The data-based monopolies of some large corporations

While this is a concern, this is already happening without the advent of Artificial Intelligence. From an AI perspective, what we really should be concerned about is that the super intelligence and singularities ALGORITHMS and MODELS are not owned and controlled by a few large non-UK corporations. We need to make sure the UK has a seat at this table and our own talent do not all work for foreign entities. More seed and growth investments for UK AI entrepreneurs is needed. Fundamentally the Algorithms and Models are the engine of machine intelligence, Data is just the fuel.

7. The ethical implications of artificial intelligence

The transparency of decision making seems to be of importance for the acceptance and adoption of AI in many areas, despite this not being the case for the equivalent human decision makers. This transparency not only relates to being able to audit the internals of the neural network in terms of the features it is using to make a decision, but also to have visibility of the data used to train the system, so that any biases can be seen, or rather it can be demonstrated that any intrinsic biases in the data have been removed or compensated for during the training process. There are other ethical concerns with AI, especially where there is not an obvious right and wrong answer and either outcome has consequences. How can we impose the ethical and moral standards of the user of the AI system rather than rely on the default programmed into the system and thus encapsulating the ethical standards of the coder or the underlying training data.

8. The role of the Government

There are a number of areas the government can support the adoption of AI and position the UK as the leading authority on AI. Firstly, ensure a strong pipeline of students from school with strong computer science teaching, to university with

more courses at undergraduate level and more secondary degree options. Secondly halting the brain drain of researchers and entrepreneur working for US or other non-UK companies. Thirdly, better support for entrepreneurs with seed and growth capital to stop startups being bought by non-UK large corporations. Fourthly, supporting initiatives like Informed.AI who are focused on knowledge sharing and education to ensure a positive perception of AI continues. Fifthly, being ready to support the universal basic salary when AI job displacement reaches levels that require the government to support the general public. Finally, embracing any minor legal changes needed to facilitate the rapid adoption of AI systems.

9. The work of other countries or international organisations.

Given my Informed.AI platform, that has users from all over the world, I get to see activity from every country. While there are the three obviously epi-centers of activity for AI, namely America, UK & China, there is one other emergent country that is noteworthy, and that is India. A hub of outsourcing IT at the moment, many of India's leading vendors are building out there AI platforms and capabilities. Informed.AI is an international organisation, with a global team and a number of global initiatives including the meet up chapters of Neurons.AI and the Global Achievement Awards for AI.

22 August 2017

Joshua Parikh – Written evidence (AIC0031)

Submission to the Select Committee on Artificial Intelligence- 28th August 2017

On Behalf of Joshua Parikh

Executive Summary: I firstly discuss formulating a definition of Artificial Intelligence. I then discuss how Artificial Intelligence and automation might have an impact on the labour market, and consider the detrimental effects which this might have. I finally consider two key frames for any policy response, and compare this to the history of policy responses which have been undertaken, suggesting a broader and more radical response is necessary.

Definition of AI and Robotics

1. Artificial Intelligence is a complex phenomenon to define, given the variety of definitions in the area- one overview paper cites 70 different definitions of intelligence¹⁰²⁸; and another article compares 5 definitions of Artificial Intelligence based on feedback from Artificial Intelligence experts¹⁰²⁹. A few takeaways are helpful: firstly, intelligence is best understood as a spectrum of capabilities, rather than a single capability. The variety of capabilities involved suggests that our evaluation of the nature of intelligence might be subject to numerous cognitive biases¹⁰³⁰.
2. Secondly, the definition of Artificial Intelligence offered by the recent Robotics and Artificial Intelligence Inquiry is inadequate. In particular, it does not have a broad enough definition of intelligence, which might include social, emotional or spiritual dimensions; nor does it have a broad enough definition of what is Artificial¹⁰³¹.

¹⁰²⁸ Legg, S., and Hutter, M., 2007, "A Collection of Definitions of Intelligence", *arXiv:0706.3639 [Cs]*, [online], Available at: <http://arxiv.org/abs/0706.3639> [Accessed 18 July 2017],

¹⁰²⁹ Faggella, Daniel, 2016, "What is Artificial Intelligence? An Informed Definition", *TechEmergence*, [online], Available at: <http://techemergence.com/what-is-artificial-intelligence/> [Accessed 18 July 2017]

¹⁰³⁰ Bostrom, N., 2014, "Superintelligence: Paths, Dangers, Strategies", Oxford: Oxford University Press, p.111

¹⁰³¹ UK Parliament, 2016, "Robotics and Artificial Intelligence: Fifth Report of Session 2016-2017", *House of Commons Science and Technology Committee*, [online], Available at: <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf> [Accessed 18 July 2017], Para 4

3. Thirdly, my own preferred definition is to “bring together the Artificial with the Intelligent”¹⁰³², with intelligence defined with Stuart Russell as “the ability to act successfully”¹⁰³³, and Artificial referring to its existence in machines, robotics, software or algorithms. Other good definitions include Legg and Hutter’s definition¹⁰³⁴, and the definition offered in the TechEmergence article¹⁰³⁵. I came to this definition through analysis of several articles, many of which are contained in the enormously helpful Global Politics of AI Reading List which has been developed by Allan Dafoe and others¹⁰³⁶.

Jobs

4. Much has been written on the impact of AI-related automation on jobs, and there is a strong disagreement about the impact of automation, because the future is uncertain and difficult to predict. I follow the analysis of Nigel Cameron in identifying the probable impact in two key ways.
5. Firstly, there is a large probability of significant labour disruption, or what Cameron describes as “substantial turbulence”¹⁰³⁷. Carl Benedikt Frey and others focus on the term “creative destruction”, following Joseph Schumpeter, arguing that “as automation makes the jobs of some workers redundant, it also creates new employment opportunities, but for a different breed of worker”¹⁰³⁸. Any process of growth therefore results in destruction of jobs which existed before, and those that benefit from any new changes may well not be the same as those who were hurt by them. Indeed, Frey et al note that “economic historians have long debated if the Industrial Revolution was worth it”,¹⁰³⁹ given the creative destruction that ensued. In more recent times, the collapse of the shipbuilding industry in the UK or other

¹⁰³² Parikh, J., 2017, “The Rise of the Machines: Preparing for the Incoming Revolution in Robotics and Artificial Intelligence”, Forthcoming

¹⁰³³ Russell, S., 2017 “Defining Intelligence”, *EDGE*, [online], Available at: https://www.edge.org/conversation/stuart_russell-defining-intelligence [Accessed 18 July 2017]

¹⁰³⁴ Legg, and Hutter, “A Collection of Definitions of Intelligence”, p.9

¹⁰³⁵ Fagella, “What is Artificial Intelligence”

¹⁰³⁶ Dafoe, A., 2017, “Reading List for the Global Politics of Artificial Intelligence”, [online], Available at: <http://www.allandafoe.com/aireadings> [Accessed 11th August 2017]

¹⁰³⁷ Cameron, N., 2017, “Will Robots Take Your Job? A Plea for Consensus”, Cambridge: Polity Press, p.11

¹⁰³⁸ Frey, C.B., Berger, T., and Chen, C., 2017, “Political Machinery: Automation Anxiety and the 2016 Presidential Election”, *Oxford Martin Programme on Technology and Employment*, [online], Available at: http://www.oxfordmartin.ox.ac.uk/downloads/academic/Political%20Machinery-Automation%20Anxiety%20and%20the%202016%20U_S_%20Presidential%20Election_230712.pdf [Accessed 11 August 2017], p.5

¹⁰³⁹ *Ibid.*, p.16

stories of industrial decline are cautionary tales, as Cameron shows. With the development of Artificial Intelligence, this disruption will take place even if one remains optimistic about job creation- we are prone to having “Rust Belts breaking out among entire sectors of the economy”¹⁰⁴⁰.

6. Secondly, there is a non-negligible probability of mass unemployment. There are several major predictions of mass automation: the OECD predicted an average of 9% of jobs would be automated across OECD countries¹⁰⁴¹; Carl Benedikt Frey and Michael Osborne, in a report produced with Deloitte, predicted that “35% of today’s jobs in the UK are at high risk of automation within the next 10 to 20 years”¹⁰⁴². Price Waterhouse Coopers predict 32% of jobs might be automated by 2032¹⁰⁴³. Many of these estimates come with heavy caveats e.g. social resistance might be sufficient to stop mass automation and/or resultant unemployment; and there are other academics who suggest this effect may not take place. Much ink has been spilled over the precise amount of automation and unemployment that is likely to result. It seems instead that “the responsible thing is to prepare for all outcomes that are seriously possible”¹⁰⁴⁴. Given that the scary estimates seem seriously possible, it is only responsible to prepare for a society where we face a collapse in the employment rate.
7. There are serious problems which might arise if either significant disruption or mass unemployment result. The first possible problem is a mass increase in material poverty- if people lose their jobs, which are historically people’s major sources of income, then we can see the probability of an increase in material poverty. Furthermore, the

¹⁰⁴⁰ Cameron, N., 2016, “Social Implications and the Future of Work”, Talk at the Faraday Institute for Science and Religion, [online], Available at: https://www.faraday.st-edmunds.cam.ac.uk/Multimedia.php?Mode=Add&ItemID=Item_Multimedia_694&width=720&height=460 [Accessed 19 July 2017]

¹⁰⁴¹ Arntz, M., Gregory, T., and Zierahn, U., 2016, “The Risk of Automation in OECD Countries: A Comparative Analysis”, *OECD Social, Employment and Migration Working Papers*, [online] Available at: http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en [Accessed 18 July 2017]

¹⁰⁴² Deloitte, 2016, “Written Evidence submitted Deloitte (ROB0019)”, [online], Available at: <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/robotics-and-artificial-intelligence/written/32514.html> [Accessed 11 August 2017]

¹⁰⁴³ PwC, 2017, “Will Robots Steal Our Jobs? The Potential Impact of Automation on the UK and Other Major Economies”, [online], Available at: <https://www.pwc.co.uk/economic-services/ukeyo/pwcukeo-section-4-automation-march-2017-v2.pdf> [Accessed 11 Aug 2017]

¹⁰⁴⁴ Cameron, “Will Robots Take Your Job”, p.86

association between joblessness and poverty is one of the highest in Europe¹⁰⁴⁵. This does not mean that a job is always sufficient to ward off poverty, particularly if costs such as housing are high. But household joblessness is not likely to help. In addition, many suggest that low-paid work is at a greater risk of automation- the report from Frey and Osborne suggests that “Across the UK, jobs paying less than £30,000 a year are nearly five times more likely to be replaced by automation than jobs paying over £100,000”¹⁰⁴⁶.

8. Insofar as we keep a “broadly global perspective”¹⁰⁴⁷, the problem might be even worse. Frey and Osborne suggest that the Developing World are likely to be hit harder by the risk of automation with the risk of 85% of jobs being automated in Ethiopia, at an extreme¹⁰⁴⁸.
9. From here, we can also evaluate the effects of worklessness on wellbeing and meaning. The remarks of sociologist William Julius Wilson are important to see how the removal of works might have broad and devastating effects across a number of areas: “the consequences of high neighbourhood joblessness are more devastating than those of high neighbourhood poverty.... Many of today’s problems in the inner-city ghetto- crime, family dissolution, welfare, low levels of social organization and so on- are fundamentally a consequence of the disappearance of work.”¹⁰⁴⁹ This is a difficult problem, suggests the inadequacy of policy interventions in a vacuum. For example, some have suggested a Universal Basic Income as a possible policy response. Insofar as this fails to deal with the crisis of meaning and more

¹⁰⁴⁵ De Graaf-Zijl, M., and Nolan, B., 2011, “Household Joblessness and Its Impact on Poverty and Deprivation in Europe”, *Gini Discussion Paper 5*, Online, Available at: http://www.gini-research.org/system/uploads/240/original/DP_5.pdf?1298997991 [Accessed 18 July 2017]

¹⁰⁴⁶ Deloitte, “Written Evidence”

¹⁰⁴⁷ Future of Humanity Institute et al, “Joint written evidence submitted by Future of Humanity Institute, Centre for the Study of Existential Risk, Global Priorities Project, and Future of Life Institute (ROB0052)”, [online], Available at: <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/robotics-and-artificial-intelligence/written/32690.html> [Accessed 11 August 2017]

¹⁰⁴⁸ Frey, C.B., and Osborne, M., 2016, “Technology At Work v2.0: The Future Is Not What It Used To Be”, *Citi GPS*, [online], Available at: http://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work_2.pdf [Accessed 11 Aug 2017]

¹⁰⁴⁹ Wilson, W.J., 1997, “When Work Disappears: The World of the New Urban Poor”, New York: Vintage

detrimental effects on wellbeing due to worklessness, it is likely to be insufficient.¹⁰⁵⁰

10. Finally, there is a strong risk of social and political unrest. Carl Benedikt Frey and others suggests that the Industrial Revolution provoked significant political unrest, and further that automation anxiety was responsible for the election of Donald Trump, which was an unprecedented measure of political disruption. They suggest that future automation might lead to similar political unrest.¹⁰⁵¹ Unrest might be spurred by related redevelopments- many have worried about whether artificial intelligence might promote greater inequality, and the Chairman of the World Economic Forum suggests that this is the “greatest societal concern”¹⁰⁵² of the Robotics Revolution, linking it to mass social unrest.

Policy response

11. It is helpful to frame the policy response to these developments. The first point is that the scale of social reform needed is very significant. The possible level of mass unemployment is extraordinarily high- for comparison, during the Great Depression, the highpoint of UK unemployment in the 20th Century, the unemployment rate reached 23%; the equivalent high in the 1980s was 14%¹⁰⁵³. Preparing for the possibility of extremely high mass unemployment or at least substantial labour disruption is a significant social problem which requires radical action, to ensure that it is not devastating.

12. A successful response to the implications of Artificial Intelligence should also be broad. Analysing the effects of AI on jobs alone has implications for a range of government policy areas, from education, with much focus on lifelong learning¹⁰⁵⁴; to the flotation of a Universal Basic Income from numerous policy theorists (a policy which requires serious

¹⁰⁵⁰ This point is made well by Brynjolfsson, E., and McAfee, A., 2014, “The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies”, London: W.V Norton and Company Ltd, p.235

¹⁰⁵¹ Frey, Berger, and Chen, “Political Machinery: Automation Anxiety and the 2016 Presidential Election”,

¹⁰⁵² Schwab, K., 2016, “The Fourth Industrial Revolution: What It Means, How to Respond”, *World Economic Forum*, [online], Available at: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> [Accessed 11 August 2017]

¹⁰⁵³ Denman, J., and McDonald, P., 1996, “Unemployment Statistics from 1881 to the Present Day”, *Office for National Statistics*

¹⁰⁵⁴ Sachs, J., 2016, “Smart Machines and the Future of Jobs”, *Boston Globe*, [online], Available at: <https://www.bostonglobe.com/opinion/2016/10/10/smart-machines-and-future-jobs/tPxRJvLpgw0W3SPrifpxTN/story.html> [Accessed 11 August 2017]

consideration but must be recognised as no “silver bullet”¹⁰⁵⁵, as economist Alison Fahey cautions, even while remaining sympathetic to UBI herself), and much elsewhere. Once this broadens out to many of the other important problems associated with AI- AI Safety, bias, security, transparency- one can see the range of implications, and the failure of single policies as a solution. Instead, a broad and coordinated strategy across multiple sectors, involving stakeholders in government, business and NGOs, is vital to ensure the most effective response to Artificial Intelligence.

13. It is also worth noting, as unlikely to be noted by other contributors to this inquiry, how it is important to involve religious groups- to focus on Christianity in the UK, the Church is an extremely powerful organisation with thousands of members who can undertake strong political engagement; a significant voice in the House of Lords with the presence of various Bishops; and a strong moral and intellectual vision which is likely to be interested by many of the questions that Artificial Intelligence raises, from the interplay with human identity to the impact on the poor- a subject the Church has previously had a strong voice on in the public sphere from benefits¹⁰⁵⁶ to payday loans¹⁰⁵⁷. Involvement from religious groups is therefore a key stakeholder to ensure a more effective response to the social implications of Artificial Intelligence.

14. These principles are worth comparing with the response to the Artificial Intelligence and Robotics Inquiry that was previously undertaken. The Inquiry, though surveying a range of important topics, made limited recommendations. The Government response to these recommendations was much weaker and failed to implement these properly, ending up only with the recommendation to put more money towards Artificial Intelligence in the Industrial Strategy, which turns out to be a much smaller investment than other nations. References to the “Government’s less than wholehearted engagement”¹⁰⁵⁸ within the Inquiry suggest that this is a deep-running problem. This is neither

¹⁰⁵⁵ Fahey, A., 2017, “Is Universal Basic Income a Viable Way to Support Humans in the Face of Technological Change?” *Effective Altruism*, [online], Available at: <https://www.youtube.com/watch?v=81JzQ55jlfQ> [Accessed 19 July 2017]

¹⁰⁵⁶ Boffey, D., 2011, “Archbishop Rowan Williams Backs Revolt Against Coalition’s Welfare Reforms”, *The Observer*, [online], Available at: <https://www.theguardian.com/politics/2011/nov/19/archbishop-rowan-williams-welfare-reforms> [Accessed 11 August 2017]

¹⁰⁵⁷ Grice, A., 2013, “War on Wonga: We’re putting you out of business, Archbishop of Canterbury Justin Welby tells payday loans company”, *The Independent*, [online], Available at: <http://www.independent.co.uk/news/uk/home-news/war-on-wonga-were-putting-you-out-of-business-archbishop-of-canterbury-justin-welby-tells-payday-8730839.html> [Accessed 11 August 2017]

¹⁰⁵⁸ “Robotics and Artificial Intelligence: Fifth Report of Session 2016-2017”, para 34

broad enough nor radical enough as a response, demonstrating a failure to properly tackle this issue and a lack of strength and stability in leadership. A better policy response will need to incorporate these principles, to ensure that the social implications of Artificial Intelligence are not catastrophic.

28 August 2017

Jonathan Penn – Written evidence (AIC0198)

September 6th 2017

Jonathan Penn, BA, MPhil, PhD (exp. 2019)

Raising, Williamson and Lipton Trust scholar at the University of Cambridge

Google/European Youth Forum Technology Policy Fellow for 2017-18

Visiting Researcher, MIT, 2018

The Role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Sankalp Bhatnagar (Center for the Future of Intelligence, The New School) and I propose that the Government pilot an AI safety certification program. The idea, which I will now outline, iterates on existing notions of algorithmic-governance (see: Rahwan, 2017; Helbing, and Pournaras, 2015). Our motivation is this: it will be difficult to win industry support globally for the regulation of AI. To ensure a sustainable future for all, regulation will need to be rolled out in stages. In our system, the first stage is a visual identification regime that allows the public to see where AI is in use, as USDA organic labels do for organic food in the United States. This early entry into the regulation arena will ensure that journalists, policy makers and regular citizens have something to point to when asked to identify, "Where is AI?" A visual marker could be used in newsfeeds, tax reports, medical search results or other value-sensitive areas in which AI (or machine learning software) is in use.

Following this, we propose that a rough taxonomy of AI be developed in partnership with experts so that different stable features can be labeled and effectively managed, as we do for art (MPAA ratings in the U.S.) and hazardous chemicals (i.e. "poisonous," "corrosive," etc.). This path would not inhibit innovation; rather it would serve as a way for the public to interact with the pace of industry. Hammond (2016) has attempted a Periodic Table of AI, for instance. A basic classification system like this could facilitate a later framework akin to an API for ethical standards (an Ethics-Programming-Interface or "EPI"). A standardized ethical protocol for developers and industry could make procedures for some ethical dilemmas open-source. Precedent in western government already exists. In 2011, the Obama Administration mandated that all U.S. agencies use web-APIs as their standard to allow public access to open-use data (via Executive Order 13571). Prior to this, each agency had to prototype and pay for its own solution. The web-API standard reduced fragmented, inefficient and costly builds. It has since been heralded as a phenomenal success; the majority of US government agencies now use it.

Labeling and classification regimes provide the public with a minimal-viable-product style pathway into the sustainable regulation of AI. Following this, we propose to explore a system in which members of the public are called upon in a manner resembling jury duty or ancient Athenian-style 'sortation' to settle AI's ethical dilemmas democratically. This jury-in-the-loop (JITL) mechanism iterates on Rahwan's notion of the "society-in-the-loop" (Rahwan, 2017).

Impact on Society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

Benjamin Franklin once summarized his view of fire safety by stating, "An ounce of prevention is worth a pound of cure." In my opinion, the majority of discussion about the future impact of AI gravitates towards a "cure" mentality rather than one of prevention. Rather than question the historical, sociological or economic systems that bring about unsustainable technological dependencies, for example, experts often focus on how to address what happens when AI tools break or malfunction and how can we prepare for the consequences. There is an implicit risk in this view that we unknowingly forfeit civic agency to industry by assuming that regulation alone will counteract those who adopt a "move fast and break things" mentality. In fact, education may be as persuasive and effective a tool as technical solutions. Take, for example, the impact of Al Gore's "An Inconvenient Truth" on shifting the public consensus around global warming. Similar efforts should be made to inform the public about what AI can do.

To foster the prevention view, we must invest in social science and humanities research that provides historical, sociological, and philosophical context for our present dilemma. We must, for instance, place the history of AI within the history of twentieth century science and technology. Many AI experts are surprised to learn, for instance, that the quasi-official histories of AI are now long outdated (McCorduck, 1979; Crevier, 1993). Those histories written by former practitioners (Simon, 1996; Newell, 2000; Boden, 2006; Nilsson, 2010) add a useful but often inward and narrow view of the field's development. The impact that military funding has had on AI research over the last sixty years has not been well studied, for example. Other histories of computing (Agar, 2006; Edwards, 1996) have shown that past technologies that are conceptually adjacent to AI such as the invention of the general-purpose computer have served to consolidate state power, silence critics, and entrench military power structures over the last two hundred years. AI, as such, cannot be treated as ahistorical. The Government should invest in new research to reverse this trend.

A reoccurring theme in my doctoral research at the University of Cambridge is how software is shaped by its user-base. When the first electronic computers were developed in the 1940-50s, for instance, lead engineers in the US and UK

adopted wildly different strategies for how to program their machines. Some systems were designed to be openly accessible, which invited participation from students and industry. Other systems were arcane, which limited use to the military or to the upper echelons of academia. The openly accessible systems led to the advent of the Open-Source Software (OSS) movement, which is widely celebrated today as a valuable public good. In relation to the future of artificial intelligence, the Government should take actions that enable the public to interact with AI as citizens and not just as consumers. To accomplish this end, investments should be made to create publically owned AI systems in Britain that solve pressing public problems like tax avoidance and wasteful energy use.

Finally, to broaden our view of AI's future, investments should be made in academic research that clarifies what the term "artificial intelligence" is meant to signify. Interpretations abound! Some academics question the extent to which AI can even be described as a coherent discipline (Bobrow and Hayes, 1984; Boden, 2006). Competing research values and the speed of innovation escalate this hurdle. Marvin Minsky, for instance, once famously deemed neural networks to be a "sterile" area of research (1969). For present and future stakeholders to reach lasting solutions and navigate fearmongering, we must build consensus around what precisely within "AI" research deserves the public's focus.

6 September 2017

PHG Foundation – Written evidence (AIC0092)

Written evidence by Sobia Raza, Alison Hall and Tanya Brigden on behalf of the PHG Foundation

Summary

We welcome this House of Lords Select Committee inquiry on Artificial Intelligence (A.I.). Whilst the field of A.I. has existed for decades, in recent years significant advances in this area have expanded the range of A.I. based applications. Our organisation is specifically interested in the use of artificial intelligence for health and healthcare and the associated opportunities, risks, ethical and social implications, and wider policy considerations. Our responses to this inquiry are therefore in the context of A.I. for health.

About the PHG Foundation

1. The PHG Foundation is an independent, not-for-profit health policy think-tank that aims to make science work for health. We provide knowledge, evidence, tools and opportunities to help policy and decision-makers put advances in the biomedical and digital health technologies within the reach of every citizen in the form of effective, affordable and more personalised healthcare. The PHG Foundation has no relevant financial or other interests to declare. Other recent consultation responses¹⁰⁵⁹ are freely available from our website along with related reports¹⁰⁶⁰, briefing notes¹⁰⁶¹ and infographics¹⁰⁶². We are happy to comment in greater depth on request, or to provide oral evidence.

The pace of technological change

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

2. Accepting there is not a universally accepted definition of A.I., in this consultation response we use the term to denote the development and use of computing systems concerned with making machines work in an intelligent way, including those that iteratively learn from data to improve their performance with experience.
3. A.I. already underpins a plethora of mainstream technologies across many life domains *e.g.* web search engines, fraud detection, marketing systems.

¹⁰⁵⁹ www.phgfoundation.org/consultations

¹⁰⁶⁰ www.phgfoundation.org/reports

¹⁰⁶¹ www.phgfoundation.org/briefing_notes

¹⁰⁶² www.phgfoundation.org/resources/infographic

Rapid developments in associated technologies and sub-areas of A.I. such as machine learning, computer vision and natural language processing, combined with the increasing availability of 'big data' are expanding the prospective applications of A.I. and transforming previously hypothetical uses into more tangible prospects; autonomous vehicles are one case in point.

4. In health and medicine there is great scope for A.I. based applications to be developed using patient health records and other health related datasets. When applied to these datasets, the view is that A.I. approaches may:
 - help to identify new disease biomarkers and refine understanding of disease
 - be used to make predictions about health and disease risk and potentially to stratify populations according to these predictions to better target appropriate interventions
 - inform and underpin new medical diagnostics, helping to develop more targeted treatments, and treat and manage patients and individuals on a more 'personalised' basis.
5. A.I. approaches are already beginning to demonstrate potential utility for very specific medical applications, examples including:
 - Automation of medical image analysis *e.g.* in radiology
 - Risk management support tools *e.g.* to identify patients at high risk of hospital readmission, or acute kidney injury¹⁰⁶³Beyond these highly targeted applications, the wider and large-scale use of A.I. in health is further from realisation due to practical, technical, and societal factors.
6. In our view, the key factors that are most likely to impact on the pace of A.I. based developments in health include:
 - Data availability for 'training' *i.e.* developing A.I. based algorithms
 - Cross-sector collaboration – particularly between the computing (A.I.) and the healthcare and medical research domains
 - The ability to collate enriched health datasets and share data within and between those sectors collaborating to develop health related A.I. applications
 - The challenge in securing public trust in sharing health data, particularly with private sector developers
 - Difficulties in predetermining user perception and preference concerning A.I. based health devices – especially where the tools interface directly with patients and the publics

¹⁰⁶³ <https://deepmind.com/applied/deepmind-health/working-nhs/how-were-helping-today/>

Is the current level of excitement which surrounds artificial intelligence warranted?

7. There is a great degree of excitement and discourse surrounding the potential impact of A.I. in health and medicine, which stems from the potential to derive new insights from health datasets. Whilst A.I. does hold great promise to benefit patients and health systems, we believe the current levels of excitement should be tempered by the immediate practical challenges and wider considerations to developing health related and medical A.I. applications, these include:
- Technical obstacles to obtaining health data sets: not least due to the slow pace of health record digitisation, but also the lack of data standards and interoperability
 - Technical challenges to collating citizen generated health-relevant data (e.g. from wearables and monitors) and integrating this with health records
 - The need for greater collaboration between A.I. experts and medical professionals in order to better define and prioritise the areas to which A.I. could be applied
 - Uncertainty surrounding the impact of upcoming regulatory changes (such as the EU General Data Protection Regulation (GDPR) coming into force) on the legitimacy of data processing and data profiling
 - Uncertainty regarding the implementation of the proposals set out by the National Data Guardian (for Health and Social Care) on 'data security and consent and 'opt-outs' recently accepted by the Government¹⁰⁶⁴, and specifically the impact of an 'opt-out' on the availability and completeness of datasets¹⁰⁶⁵, as this will influence the ability to develop and use A.I. tools which can serve a diverse U.K. population.

Impact on society

How can the general public best be prepared for more widespread use of artificial intelligence?

8. Across a range of sectors the more widespread use of A.I. is expected to impact upon the current job market, including in healthcare. Whilst some

¹⁰⁶⁴ Government response to the National Data Guardian for Health and Care's Review of Data Security, Consent and Opt-Outs and the Care Quality Commission's Review 'Safe Data, Safe Care'. July 2017.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/627493/Your_data_better_security_better_choice_better_care_government_response.pdf

¹⁰⁶⁵ PHG Foundation Consultation response to the National Data Guardian for Health and Care's Review of Data Security, Consent and Opt-Outs. September 2016.

http://www.phgfoundation.org/documents/562_1473342208.pdf

forms of health employment may be displaced by A.I. technologies, there is also the potential for new types of employment to be created, and the need for collaboration between health professionals and the A.I. sector will be increasingly important.

9. The public must therefore be prepared for an A.I. integrated healthcare workspace. This will require education and training to place greater focus on skillsets that arguably cannot easily be displaced by A.I. such as creativity, effective social interaction, manual dexterity and intelligence.
10. As the future job market may be much more fluid, support and incentives for life-long learning will be important to enable healthcare workers to acquire new skills and retrain for new work.
11. To prepare the public for the future widespread use of A.I. it is crucial to provide accessible information and ongoing engagement that highlights the existing use of A.I. in many domains of life, including its emerging use for health and healthcare. Encouraging awareness surrounding current uses of the technology may help dissolve misconceptions that fuel opposition. Early engagement, and raising awareness around the potential of A.I. to support, inform and improve healthcare, will prepare the public and health professionals for more extensive interactions with A.I. in the future.

Public perception

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

12. Since the development of A.I. applications for healthcare is contingent on data availability, public discourse around A.I. should be accompanied by greater engagement around the benefits and risks of sharing health datasets, and a concerted effort to build public trust for sharing health data.
13. Within the UK, *Understanding Patient Data*, set up following the National Data Guardian's Review of consent and opt-outs is one important initiative to support conversations with the public, patients, and healthcare professionals about the uses of health information for care. To continue to improve awareness and engagement it is crucial that such efforts are an ongoing rather than a transient programme of work.
14. The Global Alliance for Genomics and Health (GA4GH)¹⁰⁶⁶ is an organisation committed to improved genomic data sharing on a global basis particularly for medical research. It is developing a number of demonstration projects that explore the use of A.I. to facilitate effective data sharing.
15. Since the development of health based A.I. applications will require collaboration between different sectors, it will be important to embed

¹⁰⁶⁶ <http://genomicsandhealth.org/>

appropriate frameworks that can both support cross sector data sharing and also build and reinforce public trust through transparency and engagement about how health data are used.

16. In the context of healthcare, as A.I. applications develop they have great potential in the future to underpin, inform or support medical enquiries, diagnoses, health monitoring and tailored care. If integrated effectively, there is the opportunity for A.I. to not only enable greater healthcare personalisation, but also alleviate some of the current pressures on the health system. The success of these transformative technologies will in part rely upon the publics' and health professionals' willingness to use them. To realise the benefits of A.I. in health and medicine it will be crucial to encourage public and health professional engagement and provide a factual and transparent view of how developments in A.I. technologies facilitate better health.

Industry

No responses

Ethics

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

17. If the datasets used for developing (training) A.I. algorithms which underpin health applications are not sufficiently representative of the populations they are intended to serve, then it is possible the A.I. predictions may not function correctly for sections of the population underrepresented in the 'training' sets.
 - a. To mitigate against potential disparities, it will be crucial for policy makers and those developing A.I. based tools for healthcare to carefully consider population diversity when collating datasets for developing A.I. algorithms
 - b. The objective of equity should therefore be taken seriously by the sector. For example, questions about securing equitable access to research are already included as part of the NHS research ethics review process, and could be replicated within this sector.
18. It is possible that issues of liability may arise if incorrect health / medical predictions are made based on tools underpinned by A.I.
 - There is currently a lack of clarity in the literature surrounding who will be liable for errors made through use of A.I. tools. Such errors will be inevitable, especially at early stages of development. Mechanisms will need to be developed which address this problem. For example,

extension to existing NHS Indemnity¹⁰⁶⁷ could be employed (where the NHS adopts liability for negligent acts of professionals employed or owing a duty of care), or something similar adopted whereby the risks to users is shared between the manufacturer and the health service.

19. Since the A.I. approaches can reveal novel insights within datasets, mechanisms for dealing with incidental health findings may be necessary.
 - a. There is considerable debate about the extent to which use of novel technologies such as whole genome sequencing creates an ethical obligation to actively search for additional clinically actionable findings and/or to validate and treat any unsolicited incidental health findings that may arise through use of these technologies.
 - b. Similar challenges are likely to arise in the context of A.I. Thresholds for reporting potentially actionable findings will need to be identified; validation and reporting obligations evaluated; pathways clarified; and funding secured.
 - c. If these technologies are used by health care professionals, there will also be a need to assess how these technologies impact upon existing professional duties and responsibilities (both ethical and legal). If technologies are used for self-testing, then routes for further advice/action need to be clearly articulated.

In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

20. In the health sector, recent regulatory changes will necessitate increased transparency, particularly where algorithms are used for diagnosis or risk prediction. We welcome these changes to the extent that they ensure that such algorithms are used in ways that are safe and effective for patients and consumers.
21. Some A.I. algorithms are already regulated under the EU In Vitro Diagnostic Devices Directive 1998, but the scope of regulation will increase pursuant to the EU In Vitro Diagnostic Devices Regulation (2017) which comes into force in May 2022.
22. Under this Regulation, standalone medical software used for certain purposes will be regulated as IVD devices, and in order for them to be placed on the market within the EU, will have to satisfy requirements for clinical performance, performance evaluation, labelling and information provision. This will require developers and manufacturers to clearly articulate the uses for which A.I. algorithms will be put, and for the algorithms to have demonstrable clinical utility within a designated clinical population. Compliance with this Regulation is likely to be challenging for the sector.

¹⁰⁶⁷ <http://www.nhs.uk/claims/Documents/NHS%20Indemnity.pdf>

The role of government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

23. As mentioned in answer to the previous question above, the UK Government has confirmed that it will be implementing the EU IVDR and the EU GDPR since these Regulations come into force before the UK exits the EU: the EU IVDR directly regulates algorithms that are used for certain health related purposes since such algorithms are classified as *in vitro* medical diagnostic devices.
24. The EU GDPR (to be implemented in May 2022) specifically regulates profiling which is defined as automated processing of personal data for certain applications including health (GDPR Article 4(4))¹⁰⁶⁸. Article 13(2)(f) of this Regulation requires data controllers using profiling to disclose 'the existence of automated decision-making' and 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject' and Article 22 clarifies the legal bases under which such processing may be lawfully undertaken.
25. The scope of the GDPR regulates personal data (including some pseudonymised data). More detailed guidance is currently being prepared by the Information Commissioner's Office: depending on what this concludes, there may be a need for additional regulation of areas that fall outside of the EU IVDR and EU GDPR.
26. Facilitating effective governance and regulation as one of the means in which public trust and confidence can be facilitated, however empirical work on public attitudes and commercial access to data¹⁰⁶⁹ has suggested that understanding the broad uses of data, and who will be involved are seen as being even more important than ensuring that effective regulation and safeguards are in place.

Learning from others

No responses

¹⁰⁶⁸ http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

¹⁰⁶⁹ <https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf>

Contacts

Dr Sobia Raza (Head of Science)

Ms Alison Hall (Head of Humanities)

Ms Tanya Brigden (Policy Analyst, Humanities)

The PHG Foundation would be happy to respond to any further queries or clarifications

5 September 2017

Dr Andrew Philippides, Dr Paul Graham and Professor James Marshall, Professor Thomas Nowotny – Written evidence (AIC0088)

Dr Andrew Philippides, Dr Paul Graham and Professor James Marshall, Professor Thomas Nowotny – Written evidence (AIC0088)

[Submission to be found under Professor James Marshall](#)

Toby Phillips and Maciej Kuziemski – Written evidence (AIC0197)

Data as capital: inequality and power in the information economy

Submission to the House of Lords
Select Committee on Artificial Intelligence

Toby Phillips and Maciej Kuziemski
06 September 2017

We are currently public policy scholars from the Blavatnik School of Government, University of Oxford. Our combined experience spans four governments (Australia, Poland, EU, and UK) and many policy areas (technology, science, startups, industrial policy, social services, disaster recovery).

We write in our capacity as individuals, these views are our own.

1. This submission addresses questions 4 and 7 from the call for submissions:
4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?
7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

In particular, we address the asymmetries in power, choice and profit. We believe there are two key questions for the House of Lords to address: who should act on behalf of users and citizens? and, what are the potential remedies?

A new form of capital

2. It is widely believed that we're in the Fourth Industrial Revolution¹⁰⁷⁰ or the Second Machine Age¹⁰⁷¹: it is predicted that the market of data will be worth £322 billion to the UK economy, or 2.7% of GDP, by 2020¹⁰⁷². There are big wins to be had, and the potential to reap value transcends the boundaries of

¹⁰⁷⁰ Schwab, K., *The Fourth Industrial Revolution*, World Economic Forum, January 2017

¹⁰⁷¹ Brynjolfsson, E., and McAfee, A., *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W.W. Norton, April 2014

¹⁰⁷² SAS, *The Value of Big Data and the Internet of Things to the UK Economy*, February 2016

existing sectors. Almost every human activity bears the potential for data generation and collection, which will only be amplified by the rise of sophistication of data analysis techniques such as machine learning.

3. We believe data is a form of capital. Data is often described as 'the new oil'¹⁰⁷³, but such an analogy disregards the role it plays in the economy. Oil is primarily a commodity input or a consumer good. In contrast, data is essentially a capital asset¹⁰⁷⁴ - a recorded information necessary to produce another good or service. Just as any other physical asset, it can have a long term value - often extending the boundaries of time and purpose of its initial use.
4. Data is a unique form of capital for two reasons that further undermine the simple "data-as-oil" analogy:
 - 4.1. data is **non-rivalrous** - a bit of data can feed various algorithms and applications at the same time (*a barrel of oil can be used once only*)
 - 4.2. data is **non-fungible** - every bit of data (price, credit score, medical phenotype) is unique and carries different information (*a barrel of oil can substituted for an identical one*)¹⁰⁷⁵
5. For many tech companies, data is the predominant capital asset: high tech valuations don't rely on the book value of server farms and other plant, but rather in the data or information capital of users.
6. To be fair to the data-using entities, they are not the only beneficiaries. While they enjoy massive cash profits and advantages over competitors, users can enjoy novel and useful services. Every time my phone gives me navigation directions, it is providing a service using data generated by millions of other users.
7. But then we must ask: who actually owns the capital? Who has rights to it? To paraphrase John Taysom¹⁰⁷⁶: we are born with some data (name, birthdate) and acquire much more along the way, yet we have no way to access value from this asset. Meanwhile it generates value for other entities. They may offer services or discounts in exchange, and this may be a fair trade, but it is not transparent or symmetric.

¹⁰⁷³ The Economist, *The world's most valuable resource is no longer oil, but data*, 6 May 2017

¹⁰⁷⁴ *The Rise of Data Capital*, MIT Technology Review, 2016

¹⁰⁷⁵ *Ibid.*

¹⁰⁷⁶ Taysom, J. *On the importance of Data Governance with Special Reference to Finance*, in: *Connecting debates on the governance of data and its uses*, British Academy 2016

Digital serfdom

8. The world is headed towards a future — if not already there — where corporations extract extreme rents in a kind of feudal relationship with users; where users cannot meaningfully control or profit personally from their information¹⁰⁷⁷.
9. We will not bore the Lords with a history lesson about feudalism, but instead let us describe the situation today. The digital rentiers own proprietary algorithms (property). In exchange for access to critical services such as maps or search (feudal protection), digital users must allow the corporations to retain 100 per cent of profits derived from the users' personal information (an input to production, just like labour).
10. This asymmetric concentration of power, information and profit is a systemic market failure. An opportunity for a policy intervention presents itself.

Rights are ambiguous

11. It isn't always clear who owns this capital. Regulations that specify how data must be collected and protected, but not who can benefit from data use. The old saying goes "ownership is nine-tenths of the law", and current regimes do not limit a firm's ability to profit from data (aside from some restrictions on some uses of sensitive personal data).
12. There are attempts to mitigate this failure: the EU's General Data Protection Regulation (that comes into force in the UK on May 25th 2018) and a recently-announced DCMS-led Data Protection Bill give citizens more control and protection over their personal data. But stricter privacy controls will not solve the whole problem.
13. Even if ownership is clear in a technical, legal sense, this may not do much to ensure a just and equitable outcome. To return to our feudal example: most serfs had basic rights and some even entered voluntarily into serfdom to receive protection. But voluntary consent and legal rights is no defence against such asymmetry of power.
14. Even though these ambiguities need to be addressed, the rest of this submission assumes that citizens have a clear moral claim (if not a legal right) over any data that is associated with them or arises from their activities.

¹⁰⁷⁷ Fairfield, J. *Owned: Property, Privacy, and the New Digital Serfdom*, 2017

Valuable data; worthless datum

15. Despite the asymmetry in benefits (with almost all of it accruing to firms), users and citizens only have a weak claim to the value generated through the use of data. This is because of an equally immense asymmetry in capital contribution. As user's contribution may be a trivial piece of data: maybe a travel path, or a web page view, or the time taken to select another movie on Netflix. Meanwhile, firms spend millions of pounds on data science teams and physical plant, allocating massive resources in order to exploit data.
16. Each separate piece of data - each datum - has an almost negligible value. This is one of the biggest barriers to achieving fair recognition of the use of data as capital. Even if we accepted the economic rationale (data, as capital, should be compensated), any transaction costs would likely prove prohibitive. The administrative system needed to track and compensate the usage of a piece of data (say, an Uber route) is likely much more than the marginal benefit from that one piece of information.

Power to the people?

17. In solving this problem of asymmetry, we think there are two clear dimensions the Select Committee should investigate: agents and remedies.
18. The first dimension is agency: who should be empowered to act on behalf of the data rights-holder? We believe there are three levels to consider:
 - 18.1. The **rights-holder**. This is where we are today: the citizen or user is their own agent and interacts directly with the firm or entity deriving value from the data. Such an arrangement - as discussed - often leads to suboptimal outcomes where profit-maximizing firms take advantage of human irrationality and cognitive capacity.
 - 18.2. A **data trust** would be a third party acting on behalf of many users. A user could sign up to be a member of a data trust; making their choice based on philosophy, geographic coverage, membership terms, or area of expertise. Data trusts would then provide a link to negotiate between individuals, firms and the state. Trusts would mitigate the power/expertise asymmetry while addressing the trust deficit. In some sense, a data trust is nothing more than a collective bargaining agent; akin to unions (aggregating the interests of labour) or the freehold land societies of the 1800s (aggregating the interests of non-land-holders).
 - 18.3. The **state**. Representative governments are, by definition, agents of the people. They could adopt this role more forcefully by attempting to moderate the asymmetry themselves. The state could exercise this role through mandatory regulations (say, requiring big data projects to be monitored), taxes, or even partial-nationalisation.

19. The second question is about what, precisely, can be done to ameliorate the asymmetries discussed above. Without wanting to simplify a complex problem too much, we think there are three broad categories of remedy, each involving an interaction between a firm and an agent of the data rights-holder.

19.1. **Information.** At the most basic level this requires complete transparency about how and when a person's data is used, as well as some approximation of the value obtained by the firm. On the other hand, transparency is not sufficient if not supplemented by literacy and accountability efforts - a ten page explanation in incomprehensible technical jargon is of no use for a regular citizen. One promising development is the innovation of 'data receipts' - such as those developed by Digital Catapult¹⁰⁷⁸ - that provide website visitors with a plain language overview of how their personal data has been used. This would at least mean consent was fully informed, and the agent could determine with full information whether the "data for service" trade was fair.

19.2. **Control.** This would give agents fine-grained permissions over the use of data: essentially an "opt-out" button. The agent would have to agree to any usage that is not strictly required for the core service (e.g. Facebook can display my hometown on my profile, but cannot use this information to target ads). This could be coupled with a regulatory mandate that all services must have a "data-lite" option (even if it is more expensive or has less features), so that users cannot be coerced into consent. This option runs into a risk of overcomplicating the otherwise intuitive user experience of the web.

19.3. **Profit share.** Ultimately this involves compensating the rights-holder for the use of their capital. A profit share remedy may not necessarily be perfect (the costs of assigning value to each piece of data may be too hard); but rough approximations are possible. These could include data trusts receiving stocks in return for access to trust members' data, or states levying a tax in proportion to the size of an entities database.

20. Some agents are more suited to some remedies than others. For instance, if we are talking about profit sharing, it does not make sense to work at the individual level. No single user has leverage to exercise control; their data is close to worthless. Profit sharing makes the most sense when coupled with an aggregating agent, such as a data trust or the state.

21. As a final point, command and control approaches are likely to be ill-suited for data-related policies for three reasons: there are a relatively large number of actors with varying power and priorities; the state has relatively little capacity

¹⁰⁷⁸ <https://www.digitalcatapultcentre.org.uk/project/pd-receipt/>

to use traditional policy levers on cross-border entities that use an intangible asset; and when risk and uncertainty is high, good policy requires stakeholder cooperation and constant feedback. In practice, entities will have many possible routes for regulatory evasion. Any coercive measures would need to be broad and imprecise (such as a tax on major server architecture) rather than specific (such as a tax on profits derived from AI use).

22. The table below highlights the combinations that we believe are most likely to lead to useful outcomes, from green (most likely to work) to red (least likely to work).

		Remedy		
		Information	Control	Profit share
Agent	Rights-holder	Provides end users with non-actionable data usage insights	Citizens must navigate a confusing array of permissions and use-cases	Likely to be prohibitively costly to administer
	Trust	Trusts use information to become strong advocates and educators	Promise of enforceable way of representing trust members	The trust can negotiate terms for the use of member data, and then distribute proceeds
	State	The state could use this information to monitor and regulate entities	The state should not make opt-in/opt-out decisions for citizens	The state can ensure some of the data rents are directed towards public projects

6 September 2017

Professor Barbara Pierscionek and Dr John Rumbold – Written evidence (AIC0046)

Written submission for the Select Committee on Artificial Intelligence

Barbara Pierscionek

John Rumbold

Nottingham Trent University, College of Science and Technology

Our submission concerns four issues raised by Artificial Intelligence:

- 1) Privacy
- 2) Autonomy and Consent
- 3) Abuse and Misuse of Data
- 4) Accountability of AI systems

1) Privacy

There are massive privacy implications of the widespread use of artificial intelligence (AI) in our society. Artificially intelligent systems will be routinely collecting, processing, and storing vast amounts of data, much of which will be direct or indirect personal data. Smart homes will be full of devices feeding data for these AI systems to process. (Rushton 2015; Hart 2016) Smart cities will similarly be filled with monitoring devices. If these devices only processed the data for their designated purpose, this would limit the privacy implications. However, it was recently reported that the data from a robotic vacuum cleaner of domestic floor plans might be sold on. (Hern 2017) The security of many connected devices (known collectively as the Internet of Things, or IoT) is weak, due to poor design and/or the inherent limitations of the computing power on such devices. (Guinard 2015) Smart light bulbs can be “hacked” to overwhelm the controlling computer system.

Currently, artificial intelligence is very much in the development phase. Systems as diverse as Tesla cars and Amazon Echo digital assistants store large amounts of data in order to “train” these systems using real world data. These issues cannot be dealt with properly via the legal requirement for consent, given that third parties will be the subject of data processing or that the authorities could have access without consent. This storage of data has already thrown up new privacy issues. Mobile phones can have vast amounts of personal data on them, enabling such an intrusion into one’s personal life that in the USA a warrant is required to scrutinise them without consent.

Many of these issues apply more generally to the plethora of connected or non-connected monitoring devices. Insurers will ask for data from vehicles with semi-

autonomous driving modes, such as Tesla cars. Norfolk Constabulary have asked for dashcam footage of drivers using their mobile phones at the wheel.(Constabulary 2017) This initiative if replicated across the UK would involve a massive expansion of the monitoring of motorists. It will result in an asymmetry where drivers with dashcams will potentially be able to secure convictions against other drivers in situations where normally no action would be taken. They will be unlikely to volunteer such data where they are at fault. Dashcams will also be recording events around the vehicle, including off the road (for example, the Manchester bombing). This massive expansion of the network of monitoring devices in the UK bypasses the regulatory mechanisms. The people passing on their dashcam footage to law enforcement are no longer simply using it for domestic purposes. Legislation is required to ensure that rights are not eroded to the degree that we slip into a surveillance society by default

The issue of data “ownership” also requires addressing. The European Union has examined this, because of the implications for the digital economy. Clarity is arguably best achieved by legislation. The question of data ownership is particularly pertinent for wearable devices and medical implants, where the device that collects and stores the data is in the possession (or inside the body) of the person to whom the data pertains. The user ought to be able to change privacy setting on their devices in a granular way. It is not acceptable to deny services purely on the basis of privacy settings, except where data provision is an essential part of the transaction. Given the current dominance of US giants such as Facebook, Apple, Google, and Amazon, data will often be uploaded to US servers. This means that the regulation of trans-Atlantic data flows is crucial, and recent European Court of Justice rulings have highlighted the failure of consumer protection.(“Schrems v Data Protection Commissioner” 2015)

Privacy is contextual. Data subjects frequently allow different people to have access to different sets of data. When they allow one organization to have access to their data, they often do not expect them to pass it on to similar organizations. For example, the NHS has a unique collection of population-wide healthcare data that can be used to train artificial intelligence. It is important that the UK public benefit from the use of this data rather than large US digital giants.(Devlin 2017)

The issue of privacy applies particularly to technology that functions to assist – an electronic concierge or butler, if you will. A man in Arkansas accused of murder had an Echo device in his home. The police applied for a warrant to obtain any voice data from the device (although he later voluntarily permitted the electronic search)(Ortiz 2016). The age of the robotic butler will soon be upon us. Will an electronic butler be forced to divulge electronic secrets, like the robot in the film *Robot & Frank*? (Rumbold and Pierscionek 2017)

2) Autonomy and Consent

Empirical research suggests that the current model of consent for respecting privacy is inherently flawed. Privacy controls can provide false reassurance. (Brandimarte, Acquisti, and Loewenstein 2013) It has been argued that the public can be divided in three categories by their approach to privacy: 1) privacy fundamentalists 2) privacy pragmatists 3) unconcerned about privacy. (Hoofnagle and Urban 2014) Privacy fundamentalists are ideologically opposed to any form of data sharing. Those unconcerned about privacy will share their data freely. The privacy pragmatist makes a case-by-case decision on sharing data based on the benefit/harm balance. The importance of this categorization is the targeting of the privacy pragmatist by any ethico-legal framework. The greater the proportion of privacy pragmatists that can be recruited, the greater the coverage that will be achieved. The persuasion of privacy pragmatists is therefore key to the adoption and acceptance of smart homes and smart cities. This has important economic consequences. The companies that adopt the best policies should have a significant sales advantage, at least with the informed consumer. Likewise, the countries with the best ethico-legal frameworks should have a competitive advantage. This is part of the motivation for the EU Digital Single Market strategy.

3) Misuse and Abuse of Data

There needs to be robust mechanisms to prevent inadvertent and malicious release of data. This includes appropriate governance to reduce the chance of human error leading to leaks of information. The Big Data era requires a more rigorous definition of personally identifiable data, since the large amount of available data and new techniques make re-identification of previously securely anonymised data possible.

4) Accountability of AI systems

Accountability for the decisions and acts of AI systems, whether algorithms or robots, is vital for the regulation of AI and the prevention of harm related to their use. An example is the operation and use of autonomous vehicles (AVs). What safety standards should be set for these vehicles? Perfect compliance with legal requirements may increase the risk of accidents due to the lack of conformance with standard driving behaviour. For example, other vehicles may collide with the AVs when they stop unexpectedly at junctions. (Naughton 2015) If AV collision avoidance/mitigation systems prioritise passenger survival above any other consideration, this may have societal implications. There is clearly a case for government regulation.

The “right to reasons” in the General Data Protection Regulation is too weak. (Wachter, Mittelstadt, and Floridi 2017) The increasing role of algorithms poses a real threat to fairness and equality if there is not sufficient regulation and oversight. (O’Neil 2016) The mere insertion of a token human input does not alleviate the problem. (Elish 2016) The complete algorithm will be proprietary

information, but the data subject should have notification of automated decision-making and the right to know what information is inputted into the algorithm.

AI-guided systems for targeting marketing and campaigning have the ability to dramatically influence the democratic process, by presenting different realities to different subgroups of the population. This has the potential to subvert the democratic process.(Helbing et al. 2017)

References

- Brandimarte, Laura, Alessandro Acquisti, and George Loewenstein. 2013. "Misplaced Confidences: Privacy and the Control Paradox." *Social Psychological and Personality Science* 4 (3): 340–47.
- Constabulary, Norfolk. 2017. "#OpRingtone - Police Target Drivers Using Phones."
- Devlin, H. 2017. "UK Needs to Act Urgently to Secure NHS Data for British Public, Report Warns." *The Guardian*, August.
- Elish, M C. 2016. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (We Robot 2016)."
- Guinard, D. 2015. "Internet of Things: Business Must Overcome Data and Privacy Hurdles."
- Hart, J. 2016. "Smart Meter Data at Crux of Arkansas Murder Case."
- Helbing, Dirk, Bruno S Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen Van Den Hoven, Roberto V Zicari, and Andrej Zwitter. 2017. "Will Democracy Survive Big Data and Artificial Intelligence?" <https://www.scientificamerican.com/article/will-democracy-survi...>
- Hern, A. 2017. "Roomba Maker May Share Maps of Users' Homes with Google, Amazon or Apple." *The Guardian*, July.
- Hoofnagle, Chris Jay, and Jennifer M Urban. 2014. "Alan Westin's Privacy Homo Economicus." *Wake Forest Law Review* 49: 261.
- Naughton, K. 2015. "Humans Are Slamming Into Driverless Cars and Exposing a Key Flaw."
- O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.
doi:10.1057/s11369-017-0027-3.
- Ortiz, E. 2016. "Prosecutors Get Warrant for Amazon Echo Data in Arkansas Murder Case."
- Rumbold, John, and Barbara Pierscionek. 2017. "Does Your Electronic Butler Owe You a Duty of Confidentiality?" *Computer Law Review International* 18 (2).

Verlag Dr. Otto Schmidt: 48–52. doi:10.9785/cri-2017-0206.

Rushton, K. 2015. "Samsung Warns Viewers: Our Smart TVs Could Be Snooping on Your Private Conversations." *Daily Mail*.

"Schrems v Data Protection Commissioner." 2015.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7 (2): 76–99.

1 September 2017

Professor John Preston – Written evidence (AIC0014)

I write in an individual capacity, as an academic qualified in philosophy (to PhD-level) and in Artificial Intelligence (to MSc-level), and with an interest in the future of artificial intelligence. (A short version of my cv forms the final page of this document)

Please note Professor Preston's CV has been retained by the Committee, but is not included in the published submission.

My response relates only to the following questions in your call for evidence:

2. Is the current level of excitement which surrounds AI warranted?
3. How can the general public best be prepared for more widespread use of AI?
5. Should efforts be made to improve the public's understanding of, and engagement with, AI? If so, how?

For some time I've been concerned that the language in which we talk about AI (including, but of course not limited to, the term 'Artificial Intelligence' itself) promotes a misguided conception of the nature and capabilities of computational devices. That conception involves taking literally (and seriously) the idea that computational devices do some of the same sorts of things as human beings do, but that they do them faster, or more efficiently, or both. The most obvious example is *calculation*: none of us now baulk at the idea that computers and calculators perform calculations, that is, do the same sorts of things as human beings who calculate. There is no danger in this way of viewing machines when their activities are as simple as calculation. But when their activities become as complicated and crucial as the activities of AIs promise to be in the near future, this way of viewing machines ought not to go unchallenged.

I myself believe that it's misguided to think of machines in this way. (That is, I think it's a *philosophical* misconception to think that machines could think, or be intelligent. And there are arguments for this). But my evidence to your Committee depends only on the weaker supposition that it would be good to have available, and to promote to the public, some viable *alternative* way of thinking about computational devices (including AIs). That is the spirit in which my suggestion below is intended. My experience is that people who are not already fully signed up to the idea that computers are intelligent (or thinking things, or minds) find this alternative attractive and persuasive. I believe that, with promotion of the right kind, it might well gain a footing as a commonsensical, practical and popular way of thinking about computers (and AIs), and that the effect of its public dissemination would be positive.

What is the proposed alternative perspective?

In certain published articles (see my cv, below) I have defended the idea that computers can and should be thought of as what I call '*replacing-technologies*'. That is, they do not really perform the actions we think and talk of them as performing. Rather, they *replace* those actions. They make it the case that no such action is necessary. So, for example, with respect to their designated tasks, computers make it the case that no calculation (or intelligence, or thought, etc.) is involved. They are, in sum, substitutes for thought and intelligence, not examples of it. Of course, they achieve this by attaining the *goals* which we humans can typically only accomplish by use of thought or intelligence. Therein lies their usefulness (and the brilliance of their inventors). But the devices in question themselves, whether or not they come with robotic facilities, are not intelligent, or thinking. They are, if you will, our slaves (not *vice versa*).

This perspective is fleshed out in my article 'Unthinking Things' (see my cv, below). I intend to write more about this perspective, its philosophical underpinnings, its nature, and its appeal, in the near future, and would be happy to explain it to your Committee if that was of any interest to you.

My perspective and proposal suggest the following answers to the three questions I identified above:

2. Is the current level of excitement which surrounds AI warranted?

This depends entirely on exactly what has got us excited. AI *is* exciting, of course. Nothing in my perspective implies that we ought not to think of those who work in AI as undertaking intellectual work of the highest calibre. We certainly ought to recognise the brilliance of its exponents (starting with Alan Turing). But their typical *ideology* (the way they conceive of what they are doing) is not the only way of conceiving their activity. There is an alternative. And if we accept this alternative, then at least some of the excitement about AI will be seen to be based on confusion.

3. How can the general public best be prepared for more widespread use of AI?

Educating people (not just the general public, but also workers in AI) about this alternative perspective (that computers are *replacing-technologies*, rather than what we might call *achieving-technologies*) would prepare us not only for the remarkable AI projects and successes to come, but also for alternative ways in which to think and talk about these achievements, ways which should be more intelligible and intellectually and emotionally comfortable (to the general public) than the ways in which AI workers often present their achievements.

5. Should efforts be made to improve the public's understanding of, and engagement with, AI? If so, how?

Yes indeed, by presenting the public with this alternative way of thinking about what AI is, what workers are doing, etc.

17 August 2017

PricewaterhouseCoopers LLP (PwC) – Written evidence (AIC0162)

Introduction

1. This submission is made by Euan Cameron, AI leader, on behalf of PricewaterhouseCoopers LLP (PwC), the UK member firm of the PwC network. PwC is founded on a culture of partnership with a strong commercial focus, which is reflected in our stated purpose: *"to build trust in society and to solve important problems"*. We therefore feel it's our responsibility to respond to this inquiry into artificial intelligence, a technology that presents an opportunity to transform the way we live and work. To form our response we have consulted groups of practitioners and considered research commissioned by PwC and others to take advantage of our expertise and experience. This response represents the view of PwC only and not of our clients, and is intended to provide our perspectives and insights relevant to the scope of your inquiry. Euan was assisted by Rob McCargow, AI programme director, in the preparation of this report.
2. We have opted to answer questions 2, 3, 4, 6, 8, 10, and 11 from the call for evidence.

Defining AI

3. We use a broad definition of "Artificial Intelligence" (AI), a collective term for computer systems that can sense their environment, think, learn, and take action in response to what they're sensing and their objectives. AI works in four ways, defined by the level of human interaction in their processes, and the degree to which they can adapt to new situations: (1)

Table 1 - Types of artificial intelligence

	Human in the loop	No human in the loop
Hardwired / specific systems	<p><i>Assisted intelligence</i></p> <p>Helping people to perform tasks faster and better.</p>	<p><i>Automated intelligence</i></p> <p>Automation of manual/cognitive and routine/non-routine tasks</p>
Adaptive systems	<p><i>Augmented intelligence</i></p> <p>Helping people to make better decisions, with a system that</p>	<p><i>Autonomous intelligence</i></p> <p>Systems that can adapt to the situation and make decisions without human intervention</p>

learns from their interactions and
the environment

Public engagement

Is the current level of excitement which surrounds artificial intelligence warranted?

4. Our research suggests that AI could add up to \$15.7tn to the global economy in 2030, more than the current output of China and India combined and 14% higher than it would have been without the accelerating development of the technology. Of this, \$6.6tn is likely to come from increased productivity and \$9.1 trillion from consumption-side effects. (1)
5. Within this global growth, the UK GDP is expected to be up to 10.3% higher in 2030 as a result of AI – the equivalent of an additional £232bn – making it one of the biggest commercial opportunities in today’s fast-changing economy. (2)
6. In the coming years, advances in AI will impact all industries and business functions. The ultimate commercial potential is being able to do things that have never been done before, rather than simply automating or accelerating existing capabilities.
7. These gains are shared by businesses, through improved efficiency and the ability to create new intellectual property, and by society, in the improved products and services being offered (from personal assistants in mobile phones, to health care diagnoses, to improved cyber security) and by allowing time that is currently used for routine tasks to be shifted to more creative and productive activities.
8. We believe the current level of excitement surrounding AI is warranted, a view shared by high volume surveys that we have commissioned. More than 60% of the 2,500 consumers and business decision makers we surveyed in the US believe that AI can help provide solutions for many of the most important issues facing modern society, ranging from clean energy to the fight against cancer and disease. (3)

How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

9. We believe that new automation technologies will both create new jobs in the digital technology area and, through the wealth generated by productivity gains, support the creation of additional jobs in areas where tasks are less easy to automate.
10. As with any period of profound technology change, along with the economic gains of AI, there comes the potential for disruption to individuals, businesses and the state. Our research identifies around 30% of UK jobs that could be at high risk of automation by the early 2030s, lower than the US (38%) or Germany (35%), but higher than Japan (21%). This reflects, among other things, differences in the sector composition of their respective economies, and current levels of automation. (4)
11. In practice not all of these jobs may be automated for economic, legal and regulatory reasons
12. The reduction in certain roles due to automation are not evenly distributed across the workforce, with a greater proportion of jobs held by men at high risk (35%) than for women (25%).
13. Our research suggests that the factor most highly correlated with potential job automation is the education level of the worker that currently performs it: for those with GCSE-level education or lower, the estimated potential risk of automation is as high as 46% in the UK, but this falls to only around 12% for those with undergraduate degrees or higher. (4)
14. Overall, we expect the total level of employment to remain roughly constant, with the reduction in job numbers in some sectors being balanced by the creation of new jobs in others (4)
15. For individuals, a focus on adaptability will be key. The concept of a 'job for life' is now an old one - we may have to get used to the fact that the concept of 'a career for life' will also need to be modified. Ideally there should be a focus for individuals on acquiring a core set of technical and human skills, which can be adapted to a number of different roles.
16. Conventional economic wisdom generally holds that the long-term benefits of new technologies (and the new wealth thus created) outweigh the

short-term disruptive impacts on individuals and society. 19th century technology substituted human muscle, 20th century technology substituted human calculation, and 21st century technology has the potential to substitute parts of human thinking. All of these lead to increased standardisation, speed and strength. However, AI differs in that it has the potential to adapt and learn - hitherto a uniquely human capability. There remains a possibility that this could alter the economic impact of the technology in ways that make previous tech cycles less relevant analogies. However, our best estimate remains that the productivity gains and the wealth effects that follow will represent a significant net benefit to society. (4)

17. As a result, we would expect mean pre-tax incomes to rise due to the productivity gains. However, as with most periods of rapid change, these benefits may not be evenly spread across income groups, making the impact on median income less clear. (4)

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining least? How can potential disparities be mitigated?

18. Individuals stand to benefit from AI as consumers, through improved access to quality products and cheaper prices. We believe that the technology has the potential to improve lives across the socio-demographic and income spectrum, a view that our research appears to show is shared by much of the public. A survey that we undertook in the US showed that more than half of consumers believe AI will provide educational help to disadvantaged schoolchildren, and over 40% also believe AI will expand access to financial, medical, legal, and transportation services to those with lower incomes. (5)
19. Consumers also see the value in sharing their personal information for the greater good: 62% of US consumers would share their data to help relieve traffic in their cities and 57% would do so to further medical breakthroughs.
20. For consumers overall, these benefits appear to be valued more highly than their desire to protect jobs in the industry that is providing their goods or services: 69% prioritise access to more affordable and reliable transportation over preserving the jobs of taxi drivers, a trend reflected in the increasing popularity of ride-hailing apps. (5)
21. Within the current employment pool, job task composition (whether workers currently perform manual, routine or computational tasks that are more easily automated) has a degree of correlation with the educational requirements of those roles. For example, requirements are higher in the

human health and social work sector, where more than twice the proportion of employees have higher education levels (i.e. degree level or higher) than in wholesale and retail (33% compared with 15%). (4)

22. We believe that the highest potential impact of AI-driven job losses over the long-term is likely to be in the wholesale and retail trade sector, with around 2.3 million jobs at risk of automation. Manufacturing has a similar proportion of current jobs at potential risk (46%), but is a smaller employer so has lower total numbers at risk of around 1.2 million. A further 0.7 million jobs could be at potential risk of automation in human health and social work, although this represents only 17% of the jobs in the sector.

23. Whilst this will cause disruption which must be planned for, it is also important to remember that these role reductions are anticipated to take place over an extended period. Attrition and retirement will allow a proportion of this effect to occur naturally, and the role reductions will be replaced by new roles in these and other sectors (4)

24. There may be a case for some form of government intervention to ensure that the potential gains from automation are shared more widely across society through policies including: a medium term revision in the educational curriculum, to ensure our young people have as 'future-proof' a skill-set as possible; investment in vocational education, training, and retraining.

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

25. Sectors which will benefit first from AI are those where some or all of the following factors are at play: data is (or could be) extensively available; large amounts of expensive time are automatable or augmentable; a relatively small increase in the sophistication or accuracy with which the data can be processed leads to a large increase in the 'value' generated. Healthcare, automotive and financial services fall into these categories and are currently the sectors that we believe have the greatest potential for product enhancement and disruption due to AI. However, there is also significant potential for competitive advantage in particular areas of other sectors, ranging from on-demand manufacturing to individually customised entertainment and retail, to HR and recruitment. (1)

26. In healthcare those areas with the greatest potential include: a) supporting data-driven diagnosis e.g. using wearable technology to detect

small variations from the baseline in patients' health, allowing pre-emptive interventions; b) early identification of potential pandemics and tracking incidence of the disease to help prevent and contain its spread; and c) augmentation of imaging diagnostics (radiology, pathology). These benefits are driven by the magnitude of potential improvements in health outcomes and the growing availability of relevant data.

27. In automotive, the areas we have identified as the greatest potential are: a) autonomous fleets for ride-sharing; b) extension of semi-autonomous features such as driver assist; and c) engine monitoring and predictive maintenance. The potential benefits include increased utilisation and availability for both owners and users.
28. In financial services, key use cases include: a) personalised financial planning; b) fraud detection and anti-money laundering; and c) intelligent process automation of back office and customer facing operations. Consumer benefits include services that are more adaptive and personalised to the needs of the user, reduced cost and increased access to products that were previously tailored at high cost.
29. In the above industries the primary barriers to widespread adoption are user acceptance and regulatory concerns around standardisation and privacy of sensitive data.
30. Some industries are further down the adoption curve. This includes capital intensive industries with complex supply chains that would require a large degree of collaboration with third parties to make the best use of AI, such as in manufacturing and energy. Nonetheless, these industries also have potential use cases such as improved monitoring of auto-correcting and on-demand production processes, smart metering, capacity management through demand forecasting / dynamic pricing, and predictive infrastructure maintenance.

Ethics and regulation

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy."

31. Properly designed and implemented, AI will be a force for good, empowering people to achieve more and helping to tackle a number of the challenging problems faced in today's world. The risk that we must tackle is AI being allowed to operate beyond the boundaries of reasonable control. We recommend organisations adopt a reliable process of

assurance and risk control - 'Responsible AI' - that spans the 'design, test, implement and monitor' lifecycle. (6)

32. The design of any system, AI included, will require decisions of an ethical and legal nature in the prioritisation of design choices and outcomes, and how to address the experiences and needs of those who are not involved in the development of the product. These include bias, explainability, 'hard boundaries' and security.
33. There is sometimes perceived to be an expectations gap between how some organisations present their values and how they and the public believe they behave in reality. This same gap will need to be breached in AI systems, where the potential for 'information asymmetry' is greater. (7)
34. Research suggests that some existing implementations of AI produce outcomes that reflect biases in their source data, exacerbated by the lack of transparency around many 'black box' algorithms. Examples include use of proprietary AI systems that have been trialled in the US to assess a defendant's risk levels and guide court decisions on bail, sentencing and parole; and chatbots designed to mimic human interaction that have 'learned' anti-social or prejudiced behaviour after short periods of interacting with the public ('adversarial attack').
35. These have led to criticism of the ethical implications of training AI on datasets that contain inherent biases, and the lack of public scrutiny of the factors considered or prioritised by the algorithm. (8)
36. Legal issues are also raised: Article 22 of the General Data Protection Regulation (GDPR) requires the data controller to provide customers with clarity over how a decision that negatively impacts them was reached. Where AI systems are used to make important decisions such as assessing loan applications, this will require careful design and understandable systems to comply with GDPR. 'Black box' algorithms will not provide the customer with the information they are entitled to. (9) Without further explanation in such cases, consumers' confidence in AI risks being undermined.
37. Another key issue to address is the relative lack of diversity in the technology sector workforce, and in AI specifically. This has been the subject of recent public debate, with parts of the industry assessing how they can make this a more attractive and welcoming work environment for women and minority groups.
38. PwC, for example, has a 'Women in Tech' programme that seeks to address a gender gap in technology that starts in school and continues throughout careers. We have found that only 27% of female students we surveyed say they would consider a career in technology, compared to

61% of males, and only 3% of females say it is their first choice. 78% of students can't name a famous woman working in technology, compared to two thirds that can name a famous man working in technology. Addressing the reasons that women self-select out of technology roles will take an industry-wide effort to address representation. (10)

39.To address the lack of diversity in technology we have also recently announced a fully-funded technology degree apprenticeship to give more young people from a broader range of backgrounds the opportunity to get into a career in technology. This begins in September 2018 with 80 students splitting their time between study for a degree in Computer Science and work for PwC in Birmingham and Leeds. (11)

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

40.In addition to the technical and economic difficulties, the development and use of artificial intelligence poses new legal and regulatory challenges. For example: legal liability in the case of car accidents that involve software made by one company, which relied on the data from sensors made by a second company, to operate a vehicle assembled by a third company, owned by an end user who may have an obligation to take manual control of the vehicle if the computer demands it. (4)

41.As AI technology develops, the Government may wish to consider a number of issues in relation to regulation. This could include: regulation of the algorithms themselves in specific use cases; the processes, people and engineers, used to build AI driven systems; or the final product that contains the AI (eg automobile, medical device etc.). A variety of approaches are taken in other industries, with qualifications, standards and laws developed and enforced by a range of regulatory and trade bodies. In most safety critical industries a regulatory framework already exist and there may be the potential to apply or adapt these for AI.

42.Whether a more extensive practitioner regulation is required, or a version of the 'hippocratic oath' (13), is a matter of continuing debate in which the need for safety and security must be balanced with the need to foster innovation. The study of analogous industries may be informative in this respect.

43.It is possible that the economic benefits of AI are unevenly skewed towards those with the skills to adapt to an increasingly digital economy, placing a premium on education both before entering the workplace and when the need to reskill arises. Two in five people surveyed in the UK are worried that automation is putting their job at risk and 46% believe that

governments should take any action needed to protect jobs from automation (12).

44. A range of potential policy responses to this situation should be considered. One priority could be working with employers and education providers, to help guide investment in the most effective types of education and vocational training
45. Central and local government bodies could also consider the option of developing a framework of support for digital sectors and associated job creation opportunities. For example through place-based strategies centred around university research centres, science parks and other enablers of business growth. This place-based approach is one of the key themes in the government's new industrial strategy and its wider devolution agenda, which involves extending digital infrastructure beyond the major urban centres to facilitate small digital start-ups in other parts of the country.
46. Consideration should also be given to how the UK's position as a European destination of choice for high-tech skills should be protected and enhanced. (4)

What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

47. In our recommendations to the G20 on creating the conditions for emerging technologies to benefit people and the planet, we focused on data, algorithms, ethics, and people. All are applicable to the UK specifically as well as the wider G20: (8)
 - a. Support the creation of a better data environment – including access and data skills – to maximise the opportunity of big data and machine learning for sustainable solutions;
 - b. Develop a policy framework that supports tech companies, research institutes and universities to manage potential systemic bias in algorithms;
 - c. Consider and evaluate ethical aspects of the relationship between people and machine systems, which would include implications for privacy, scope and boundaries to human / digital augmentation and the rights of people;
 - d. Recognise and support the work being done to give every person in the world a unique digital identity.

References

1. PwC. Sizing the prize - what's the real value of AI for your business and how can you capitalise? [Online] 2017. <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>
2. The economic impact of artificial intelligence on the UK economy. [Online] June 2017. <http://www.pwc.co.uk/economic-services/assets/ai-uk-report-v2.pdf>
3. Bot me: A revolutionary partnership: How AI is pushing man and machine closer together. [Online] April 2017. <https://www.pwc.com/us/en/press-releases/assets/img/bot-me.pdf>
4. UK Economic Outlook. [Online] March 2017. <https://www.pwc.co.uk/economic-services/ukeyo/pwc-uk-economic-outlook-full-report-march-2017-v2.pdf>
5. PwC US. Bot.Me: A revolutionary partnership - How AI is pushing man and machine closer together. [Online] <http://pwcartificialintelligence.com/>
6. PwC. Accelerating innovation: how to build trust and confidence in AI. [Online] 2017. <http://www.pwc.co.uk/audit-assurance/assets/pdf/responsible-artificial-intelligence.pdf>
7. Tracey Groves, PwC partner. The Impact of AI on Ethics. MCA. [Online] 30 June 2017. <https://www.mca.org.uk/news/updates/the-impact-of-ai-on-ethics>
8. PwC. Enabling a sustainable Fourth Industrial Revolution: How G20 countries can create the conditions for emerging technologies to benefit people and the planet. [Online] 31 July 2017. http://www.g20-insights.org/policy_briefs/enabling-sustainable-fourth-industrial-revolution-g20-countries-can-create-conditions-emerging-technologies-benefit-people-planet/
9. Zoldi, Scott, Jennings, Andrew and Kinch, Brian. GDPR and Other Regulations Demand Explainable AI. FICO Blog - Analytics and Optimisation. [Online] 24 May 2017. <http://www.fico.com/en/blogs/analytics-optimization/gdpr-and-other-regulations-demand-explainable-ai/>
10. PwC. Women in Tech: Time to Close the Gender Gap. [Online] 2017. <http://www.pwc.co.uk/who-we-are/women-in-technology/time-to-close-the-gender-gap.html>
11. Innovative new PwC tech degree apprenticeship launched to address the UK's future skills gap. [Online] 12 June 2017. <https://www.pwc.co.uk/press-room/press-releases/innovative-new-pwc-tech-degree-apprenticeship-launched-to-address-the-uks-future-skills-gap.html>

12. PwC. Workforce of the Future. [Online] 25 August 2017

<https://www.pwc.co.uk/press-room/press-releases/UK-workers-ready-to-reskill-to-tackle-technology-impact-on-jobs.html>

13. The RSA. A Hippocratic Oath for AI Developers? It May Only Be A Matter Of Time. [Online] 13 February 2017

<https://www.thersa.org/discover/publications-and-articles/rsa-blogs/2017/02/a-hippocratic-oath-for-ai-developers-it-may-only-be-a-matter-of-time>

6 September 2017

Privacy International – Written evidence (AIC0207)

September 6, 2017

Frederike Kaltheuner, Programme Lead and Policy Officer, Privacy International

Dana Polatin-Reuben, Technology Officer, Privacy International

Statement of interest

Privacy International welcomes the opportunity to respond to this inquiry by the House of Lords Select Committee on Artificial Intelligence ('AI'). Privacy international is a non-profit, non-governmental organization based in London, dedicated to defending the right to privacy around the world. Established in 1990, Privacy International undertakes research and investigations into government surveillance and data exploitation in the private sector with a focus on the technologies that enable these practices. To ensure universal respect for the right to privacy, Privacy International advocates for strong national, regional and international laws that protect privacy around the world. It has litigated or intervened in cases implicating the right to privacy in the courts of the United States, the UK, and Europe, including the European Court of Human Rights and the European Court of Justice. It also strengthens the capacity of partner organizations in developing countries to identify and defend against threats to privacy. Privacy International employs technologists, investigators, policy and advocacy experts, and lawyers, who work together to understand the technical underpinnings of emerging technology and to consider how existing legal definitions and frameworks map onto such technology.

The pace of technological change

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1. Artificial intelligence (AI), or intelligent systems which can act without being specifically programmed to follow certain steps or instructions¹⁰⁷⁹, is a term that is often used to refer to a diverse range of applications and use-cases at different levels of complexity and abstraction. The term is employed to encompass everything from machine learning which makes inferences, predictions, and decisions about individuals, and other domain-specific AI algorithms, to fully autonomous and connected objects, as well as the futuristic idea of Singularity. This lack of definitional clarity is a challenge, since different types of AI and different domains of application raise specific ethical and regulatory issues.

¹⁰⁷⁹ Negnevitsky, M., 2005. *Artificial intelligence: a guide to intelligent systems*. Pearson Education.

2. The most widespread AI methods are collectively known as machine learning, which undergirds everything from text auto correction to drone targeting systems. Machine learning uses algorithms trained with vast amounts of data to improve a system's performance at a task over time. Tasks often involve making decisions or recognising patterns, with many different possible outputs in a range of domains and applications.
3. As an organisation which works on the right to privacy, we are primarily concerned about current and future applications of AI that are designed for the following purposes: (1) to identify and track individuals; (2) to predict or evaluate individuals or groups and their behaviour; (3) to automatically make or feed into consequential decisions about people or their environment; and (4) to generate, collect and share data.
4. **AI applications can be used to identify and thereby track individuals across different devices, in their homes, at work and in public spaces.** For example, while personally identifiable information (PII) is routinely anonymised within datasets, AI can be employed to de-anonymise this data, complicating the distinction between PII and non-PII data on which current data protection regulation is based.
5. **Using machine learning methods, highly sensitive information can also be inferred, or predicted from non-sensitive forms of data.** As a result of such profiling, databases that merely contain data about an individual's behaviour can be used to generate unknown data about their likely identity, attributes, interests, or demographic information. Such predictions may include information about health, political opinions, sexual orientation, or family life.
6. **AI systems can be used to make or inform consequential decisions about people or their environment.** Automated decision-making that relies on AI also plays a role in the personalisation of information and experiences, from news feeds to targeted advertising and recommendation systems. Such personalisation gears information towards individuals' presumed interests or identities, which are derived through profiling.
7. AI-driven consumer products, from smart home appliances to phone applications are often built for data exploitation. **Consumers are commonly faced with an informational asymmetry as to what kinds of data and how much data their devices, networks, and platforms generate, process, or share.** As we bring ever more smart and connected devices into our homes, workplaces, public spaces and onto our bodies, educating the public about such data exploitation becomes pressing.

8. **These applications of AI have the potential to undermine fundamental rights and liberties, from the right to privacy, freedom of expression and assembly, and raise very serious concerns surrounding discrimination.**
9. **They also have the potential to transform society as we know it.** Today, AI CCTV security systems can classify people, follow them through a crowd and detect 'suspicious behaviour'¹⁰⁸⁰; tomorrow, CCTV cameras and drones may be able to transcribe conversations through lip reading.¹⁰⁸¹ Today, insurance companies analyse how many exclamation points we use in social media posts to determine whether we are a safe driver¹⁰⁸²; tomorrow, marketers could assess our credit worthiness from objects and facial expressions in the pictures we share on social media platforms.

Is the current level of excitement which surrounds artificial intelligence warranted?

10. AI, if implemented responsibly, can have many exciting impacts on society. AI systems could improve crop yield in large-scale farming by tracking potential issues such as pests¹⁰⁸³, and interactive robots are already improving the social skills of people on the autism spectrum¹⁰⁸⁴.
11. Privacy International is not against the use of artificial intelligence; however, as is the case with most emerging technologies, there is a very real risk that commercial and government uses of AI fall into the trap of technological solutionism – the urge to fix problems that don't exist, or for which there is no technological solution, or for which a technological solution will exacerbate existing problems and fail to address underlying issues.

Impact on society

¹⁰⁸⁰ Toomey, M., 2017, Hitachi built an AI security system that follows you through a crowd. *Quartz*. Available from: <https://qz.com/958467/hitachi-built-an-ai-security-system-that-follows-you-through-a-crowd/>. [Accessed 1st August 2017]

¹⁰⁸¹ Morgan, T., 2016, Lip-reading technology breakthrough to be used on CCTV. *The Telegraph*. Available from: <http://www.telegraph.co.uk/news/2016/03/25/lip-reading-technology-breakthrough-to-be-used-on-cctv/>. [Accessed 1st August 2017]

¹⁰⁸² Ruddick, G., 2016, Admiral to price car insurance based on Facebook posts. *The Guardian*. Available from: <https://www.theguardian.com/technology/2016/nov/02/admiral-to-price-car-insurance-based-on-facebook-posts>. [Accessed 1st August 2017]

¹⁰⁸³ McFarland, M., 2017, Farmers turn to artificial intelligence to grow better crops. *CNN*. Available from: <http://money.cnn.com/2017/07/26/technology/future/farming-ai-tomatoes/index.html>. [Accessed 1st August 2017]

¹⁰⁸⁴ Available from: <https://robots4autism.com>. [Accessed 1st August 2017]

How can the general public best be prepared for more widespread use of artificial intelligence?

12. **Novel applications and recent advances in artificial intelligence could negatively affect the right to privacy. This is significant since privacy is the lynchpin of indispensable individual values** such as human dignity, personal autonomy, freedom of expression, freedom of association, and freedom of choice,¹⁰⁸⁵ as well as broader societal norms.¹⁰⁸⁶
13. The privacy implications of AI stem from its ability to recognise patterns and increasingly “derive the intimate from the available”¹⁰⁸⁷. AI methods are being used to identify people who wish to remain anonymous; infer and generate sensitive information about people from their non-sensitive data; profile people based upon population-scale data; and make consequential decisions using this data which profoundly affect people’s lives.
14. For instance, machine learning systems have been able to identify about 69% of protesters who are wearing caps and scarfs to cover their face.¹⁰⁸⁸ FindFace, a Russian face recognition application launched in early 2016, allows users to photograph people in a crowd and compares their picture to profile pictures on the popular social network VKontakte, identifying their online profile with 70% reliability.¹⁰⁸⁹ The technology has also been used to identify the real names of sex workers in adult films.¹⁰⁹⁰

¹⁰⁸⁵ Payton, T. and Claypoole, T., 2014. *Privacy in the age of Big data: Recognizing threats, defending your rights, and protecting your family*. Rowman & Littlefield.

¹⁰⁸⁶ Post, R.C., 1989. The social foundations of privacy: Community and self in the common law tort. *California Law Review*, pp.957-1010. Summarizing Post see Doyle, T., 2012. Daniel J. Solove, Nothing to Hide: The False Tradeoff between Privacy and Security. (“As the legal theorist Robert Post has argued, privacy is not merely a set of restraints on society’s rules and norms. Instead, privacy constitutes a society’s attempt to promote civility. Society protects privacy as a means of enforcing order in the community. Privacy isn’t the trumpeting of the individual against society’s interests but the protection of the individual based on society’s own norms and values”).

¹⁰⁸⁷ Calo, R., 2017. Artificial Intelligence Policy: A Roadmap. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3015350

¹⁰⁸⁸ Singh, A., Patil, D., Reddy, G.M. and Omkar, S.N., 2017. Disguised Face Identification (DFI) with Facial KeyPoints using Spatial Fusion Convolutional Network. *arXiv preprint arXiv:1708.09317*. ACM. <https://arxiv.org/pdf/1708.09317.pdf>

¹⁰⁸⁹ Available from: <https://www.theguardian.com/technology/2016/may/17/findface-face-recognition-app-end-public-anonymity-vkontakte> . [Accessed 1st August 2017]

¹⁰⁹⁰ Available from: <http://www.newsweek.com/porn-actress-facial-recognition-findface-sex-worker-453357>. [Accessed 1st August 2017]

15. A 2015 study by researchers at the French Institute for Research in Computer Science showed that 75% of mobile phone users can be re-identified within a dataset using machine learning methods and just two smartphone apps, with the probability rising to 95% if four apps are used.¹⁰⁹¹
16. Emotional states, such as confidence, nervousness, sadness, and tiredness, for instance, can be predicted from typing patterns on a computer keyboard.¹⁰⁹² The Big-Five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience) can be predicted from standard mobile phone logs.¹⁰⁹³ In 2012, Cambridge researchers used predictive modelling to analyse a dataset of Facebook Likes, demographic profiles, and psychometric tests from 58,000 Americans. From the Likes data, the model could discriminate between heterosexual and homosexual men in 88% of cases; African Americans and Caucasian Americans in 95% of cases; and Democrats and Republicans in 85% of cases.¹⁰⁹⁴
17. **While such profiling using machine learning can be highly privacy invasive, there is also no guarantee that the profile that is created in the process is even accurate, given that machine learning methods are inherently probabilistic.** Poor quality data, or systematically biased data are a common concern. Yet, even if profiling was based on perfect data, individuals could still be misclassified, misidentified or misjudged, and such errors may disproportionately affect certain groups of people (see our response to the next question).
18. Profiling, whether it relies on complex machine learning or more straightforward methods, merely determines that an individual is *highly likely* to be female, *likely* to be unworthy or credit, or *unlikely* to be married, homosexual or an introvert. **Since individuals are often unaware about the fact that they are being profiled, it can be difficult to challenge or correct inaccurately inferred or predicted information.** Do we want to rely on probabilistic knowledge to make

¹⁰⁹¹ Achara, J.P., Acs, G. and Castelluccia, C., 2015, October. On the unicity of smartphone applications. In *Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society* (pp. 27-36). ACM. <https://arxiv.org/pdf/1507.07851v2.pdf>

¹⁰⁹² Epp, C., Lippold, M. and Mandryk, R.L., 2011, May. Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 715-724). ACM. <http://hci.usask.ca/uploads/203-p715-epp.pdf>

¹⁰⁹³ de Montjoye, Y.A., Quoidbach, J., Robic, F. and Pentland, A., 2013, April. Predicting Personality Using Novel Mobile Phone-Based Metrics. In *SBP* (pp. 48-55). <https://link.springer.com/content/pdf/10.1007/978-3-642-37210-0.pdf#page=63>

¹⁰⁹⁴ Kosinski, M., Stillwell, D. and Graepel, T., 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), pp.5802-5805. <http://www.pnas.org/content/110/15/5802.full#F1>

decisions about life or death? And do we feel comfortable using uncertain and possibly discriminatory inferences to limit an individual's freedom?

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

19. AI's benefits and harms are currently distributed unequally.

Industry gains most from AI, with large tech companies (and selected government agencies) having unprecedented access to vast troves of data on billions of people around the world. Consumers and citizens are frequently unaware about the scope, granularity, and sensitivity of data that third parties hold about them, or that their data is being used to train and develop AI systems.

20. Risky applications of AI often disproportionately affect those that are already most vulnerable in society. A good example is AI-driven automated decision-making in hiring. Highly skilled job seekers have the ability to demonstrate their skills and character in a personal interview, while the low-wage sector with high turnover increasingly relies on automated and often proprietary and opaque hiring software that may rely on poor quality or inaccurate data and produce biased, inaccurate, discriminatory or unfair decisions. Such selective reliance on AI-driven decision-making is also evident in policing. While predictive policing is becoming increasingly common in UK law enforcement, it is predominantly used to fight street-level crime, rather than white collar crime such as tax evasion or fraud.

21. Finally, AI systems can contribute to the perpetuation of existing injustices and inequalities in society through inbuilt bias and discrimination. In the United States, risk assessment software purporting to predict the likelihood of reoffending has been used to aid sentencing decisions since the early 2000s. A 2016 study by the non-profit news organisation ProPublica revealed this software's bias against African-Americans, who are more likely to be given a higher risk score compared with white offenders charged with similar crimes.¹⁰⁹⁵ Another important case is facial recognition software. The US House Committee on Oversight and Government Reform found that the FBI facial recognition database contains photos of half of US adults without consent, and the

¹⁰⁹⁵ Angwin, J., Larson, J., Mattu, S and Kirchner, L., 2016, Machine Bias. *ProPublica*. Available from: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. [Accessed 1st August 2017]

algorithm is not only wrong nearly 15% of time, but is also more likely to misidentify black people.¹⁰⁹⁶

22. Machine learning can unintentionally, indirectly, and often unknowingly recreate discrimination from past data. Since profiling using machine learning can create uncannily personal insights, there is a risk of it being used against those who are already marginalised. Even if data controllers take measures to avoid sensitive attributes in automated processing, trivial information can correlate with sensitive information, potentially leading to illegal but indirect discrimination.²⁸ In racially segregated cities, for instance, postcodes may be a proxy for race. Therefore, without explicitly identifying a data subject's race, profiling may nonetheless identify attributes, or other information that would lead to discriminatory outcomes, if they were to be used to inform or make a decision.
23. Machine learning can also lead to "rational discrimination" – when data analysis finds an accurate correlation, that society nonetheless would consider discriminatory. An example would be if an algorithm found that men are less reliable in paying back loans, and hence their interest rate should be higher. Would we want to discriminate based on gender? And finally, there is simply unfairness, which might not be illegal, but could nonetheless be seen as unfair. If, for instance, a hiring software based on machine learning concludes that users of Internet Explorer are less qualified candidates¹⁰⁹⁷, we could consider this unfair.

Public Perception

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

24. **In the public imagination, AI is always something that isn't quite there yet, that is embodied, futuristic, but not yet widespread. This need to change.** This misconception risks steering the focus of regulatory discussions on speculative technologies that have yet to be implemented on a mass scale, if at all.
25. In particular, we find that **the public's understanding of AI to identify individuals across devices and in public space, and to gain highly**

¹⁰⁹⁶ See <https://oversight.house.gov/newsarticle/facial-recognition-database-used-fbi-control-house-committee-hears/>

¹⁰⁹⁷ 2013. Robot Recruiters. *The Economist*. Available from: <https://www.economist.com/news/business/21575820-how-software-helps-firms-hire-workers-more-efficiently-robot-recruiters>. [Accessed 1st August 2017]

sensitive insights from everyday traces of data, is low.¹⁰⁹⁸ Since *informed* consent is one legal ground for the processing of personal data, this lack of understanding raises concerns.

26. **Similar challenges apply to the privacy and security risks of AI-driven consumer products.** A good example is iRobot, the Roomba robotic vacuum. The product's chief executive suggested that the company might begin selling floor plans of customer's homes, derived from the movement of their autonomous cleaner, to Amazon, Apple, and Google Alphabet.¹⁰⁹⁹
27. Relevant actors, including the government, the EU Commission [the European Data Protection Board], supervisory authorities and civil society must design and develop a plan to educate data subjects and consumers about the various ways in which their data is being used by data controllers. They must also be made aware of how to gain information about processing of their data, how to exercise their rights in relation to such processing, and how to obtain redress, which requires effective implementation and enforcement of the rights of data subjects as set out in the upcoming General Data Protection Regulation (GDPR) and any related U.K. legislation.

Ethics

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

28. **AI-driven applications sort, score, categorise, assess, and rank people, often without their knowledge or consent.** We have already mentioned the privacy implications of this, but it is important to stress that other human rights are affected as well. This view is echoed by the United Nations Human Rights Council, which on 22 March 2017 noted with concern "that automatic processing of personal data for individual profiling may lead to discrimination or decisions that otherwise have the potential to affect the enjoyment of human rights, including economic, social and cultural rights".¹¹⁰⁰

¹⁰⁹⁸ The Royal Society, 2017, Machine learning: the power and promise of computers that learn by example. *Royal Society*. Available from <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>. [Accessed 1st August 2017]

¹⁰⁹⁹ Hern, A., 2017, Roomba maker may share maps of users' homes with Google, Amazon or Apple. *The Guardian*. Available from: <https://www.theguardian.com/technology/2017/jul/25/roomba-maker-could-share-maps-users-homes-google-amazon-apple-irobot-robot-vacuum>. [Accessed 1st August 2017]

¹¹⁰⁰ U.N. Human Rights Council Resolution on the Right to Privacy in the Digital Age, U.N. Doc. A/HRC/34/7, 23 Mar. 2017, para.2

1.

29. When all data, from how we fill out a form, to our location data can be used to gain even more intimate details about our lives and make consequential decisions, from access to credit and insurance to policing, this might result in widespread chilling effects. Individuals might pre-emptively self-censor their speech and behaviour, if the data it generates might be used against them.

30. AI also plays a role in personalisation of information, products, and experiences. By excluding content deemed irrelevant or contradictory to the user's beliefs or presumed interests, such forms of personalisation may reduce the diversity of information users encounter.¹¹⁰¹ Personalisation of not just information but also our perception of the world around us will become increasingly important as we move towards connected spaces, like smart cities, as well as augmented and virtual reality. An environment that knows your preferences and adapts itself according to these presumed interests would be highly personalised, but would also raise important questions around autonomy and the ethics of subtle manipulation.

In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

31. We would like to draw the Committee's attention to the work of Jenna Burrell¹¹⁰², who distinguishes between three forms of opacity: (1) opacity as intentional corporate or state secrecy; (2) opacity as technical illiteracy; and (3) an opacity that arises from the characteristics of machine learning algorithms and the scale required to apply them usefully. Only the latter implies that the system's outcomes are not be predictable by its designer, whereas users, regulators or the general public will find all the instances opaque. We are most concerned about highly complex AI systems have the potential to produce potentially harmful or dangerous outcomes that are *neither predictable by their designer nor easily discoverable by the public*.

32. **Black boxing should not be permissible wherever AI systems are used to make or inform consequential decisions about individuals or their environment;** in such instances, a lack of transparency is highly problematic. Consequential decisions are decisions that produce irreversible effects, or effects that can significantly affect an individual's life or infringe on their fundamental and human rights.

¹¹⁰¹ Pariser, E., 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK.

¹¹⁰² Jenna Burrell, *supra* note 7.

33. We would like to stress that we **are equally concerned about opaque AI systems that automatically make and those that inform decisions** that is decisions that are formally attributed to humans but are *de facto* determined by an opaque AI system. A good example is the use of automated risk scores in the criminal justice system. Proprietary software, such as the COMPAS risk assessment that was sanctioned by the Wisconsin Supreme Court in 2016¹¹⁰³, calculates a score predicting the likelihood of committing a future crime. Even if final decisions are made by a judge, the software's automated decisions can be decisive, especially if judges rely on them exclusively or haven't been warned about their risks, including that the software produced inaccurate, illegal, discriminatory, or unfair decisions.
34. **It is also crucial to define what kind of remedies different stakeholders require.** Individuals should be provided with sufficient information to enable them to fully comprehend the scope, nature, and application of AI, in particular with regards to what kinds of data these systems generate, collect, process, and share. When AI algorithms are used to generate insights or make decisions about individuals, users as well as regulators should be able to determine how a decision has been made, and whether the regular use of these systems violates existing laws, particularly regarding discrimination, privacy, and data protection. Governments and corporations who rely on AI should publish, at a very minimum, aggregate information of the kind of systems being developed and deployed.¹¹⁰⁴

The role of the Government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

35. The question of whether artificial intelligence *can* or *should* be regulated is complicated by the fact that artificial intelligence lacks a stable, consensus definition or instantiation.¹¹⁰⁵ Furthermore, an identical AI application can raise different regulatory and ethical concerns, depending on the domains in which it is employed.

¹¹⁰³ Citron, D., 2016, (Un)Fairness of Risk Scores in Criminal Sentencing. *Forbes*. Available from: <https://www.forbes.com/sites/daniellecitron/2016/07/13/unfairness-of-risk-scores-in-criminal-sentencing/#6074794b4ad2>. [Accessed 1st August 2017]

¹¹⁰⁴ Cf. Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, U.N. Doc. A/HRC/23/40, paras. 91-92 (17 April 2013).

¹¹⁰⁵ Calo, R., 2017. Artificial Intelligence Policy: A Roadmap. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3015350

36. Take for instance "SKYNET", a programme by the United States National Security Agency (NSA) which reportedly collects in bulk the metadata communication of the entire Pakistani mobile phone network, and then uses a random forest machine learning algorithm to rate "each person's likelihood of being a terrorist".¹¹⁰⁶ The insights and classifications that machine learning generates are inherently probabilistic – there are always false positive and false negatives. But the implications of this are vastly different, depending on where exactly machine learning is being applied. An exceptionally low false positive rate is remarkable in business applications, such as targeted advertising. In the case of government surveillance, however, even an error rate as low as "0.008 percent of the Pakistani population" still corresponds to 15,000 people potentially being misclassified as "terrorists".¹¹⁰⁷
- 37. What clearly should be regulated is the following: the data that feeds into AI systems; the data (and insights) that AI systems generate; as well as how and whether AI systems should be used to make or inform consequential decisions about individuals and groups, especially if these systems are highly complex and opaque.**
38. While data is central to the development of AI, in particular machine learning, governments and regulators have a responsibility to ensure that the current excitement about AI does not become a pretext for exploiting people's data without their knowledge or unambiguous and informed consent, for processing purposes that are often unexpected and may result in tangible harm.¹¹⁰⁸ We would like to draw the Committee's attention to principles such as "data minimisation", "privacy and security by design", as well as "purpose limitation" that are designed to mitigate the power imbalance between data controllers and data subjects.
39. The upcoming GDPR contains provisions that specifically address profiling and automated individual decision-making. These are necessary but not sufficient to address all privacy concerns of AI. However, a number of viable expressions in the GDPR are unclear or ambiguous, which may lead to confusion, enforcement gaps or asymmetries. We encourage the government to support additional guidance that clarifies ambiguous terms in a way that guarantees the strongest protections for data subjects.

¹¹⁰⁶ For more information, see Cole, D., 2014. We kill people based on metadata. *The New York Review of Books*, 10, p .2014.; Grothoff, C. and Porup, J., 2016. The NSA's SKYNET program may be killing thousands of innocent people. *Ars Technica.*, available at <https://arstechnica.co.uk/security/2016/02/the-nasas-skynet-program-may-be-killing-thousands-of-innocent-people/>.

¹¹⁰⁷ *ibid.*

¹¹⁰⁸ See for instance Hill, K., 2016, This sex toy tells the manufacturer every time you use it. *Fusion*. Available from: <http://fusion.kinja.com/this-sex-toy-tells-the-manufacturer-every-time-you-use-1793861000>. [Accessed 1st August 2017]

40. We strongly believe that civil society organisations should be able to investigate and lodge complaints independently or on behalf of data subjects if processing is unlawful. There is urgent need for clear EU-wide guidelines on how to claim redress in front of national supervisory authorities or national courts for violations of their rights in relation to profiling, AI, and the use of machine-learning algorithms.

8 September 2017

Raymond Williams Foundation – Written evidence (AIC0122)

Preface

The Raymond Williams Foundation, upon whose behalf, this evidence is submitted, is a charity committed to liberal adult education. Raymond Williams stated, "I've often defined my own social purpose as the creation of an educated and participating democracy". As part of our work, we organise residential and other courses, including for disadvantaged students who we support through grants and bursaries.

Specific to this submission of evidence, we also support a nationwide network of community-based discussion groups, centred on the premise of self-education via structured discussion. Some of our discussion groups have addressed issues germane to the Select Committee's current consultation, over the last few years. As a result, we therefore issued an invitation to our participants to consider the Select Committee's consultation.

The responses have been compiled into this document. It is submitted to the Select Committee on behalf of the Raymond Williams Foundation.

David Whalley
on behalf of
The Raymond Williams Foundation
<http://www.raymondwilliamsfoundation.org.uk>

Some Definitions

We consider the term **artificial intelligence** (AI) to mean intelligence exhibited by machines, as opposed to intelligence exhibited by humans or other animal species. Such an intelligent entity is taken to mean any machine or device that perceives its surroundings and autonomously takes actions that maximize its chance of success at a predefined goal. It is often cynically stated that, "AI is what no machine can do yet." Or, as machine capabilities increase, it is tempting to define natural intelligence as that which no machine can yet achieve. We do not consider that either of these last two interpretations will assist the Select Committee.

We consider that artificial intelligence is closely associated with **machine learning**. We consider the term machine learning to mean the ability of a machine or device to learn without being specifically programmed. Programming, too, has often been referred to cynically by the phrase, "Garbage

in; garbage out.” It is most important to understand that digital machines long ago passed from the world of being dependent on explicit programming.

Machines that exhibit machine learning are doing exactly that: building up their own impressions of, and interactions with, their surroundings. In order to do this, they acquire and then process statistically vast amounts of data. For example, the ability of a machine to learn language translation skills does not depend on the relevant rules of grammar, but depends on treating every word as a potential exception and analysing statistically vast bodies of related text at vast speeds.

Our Responses to your Questions

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1 We are not competent to express a view on this question.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

2.1 The excitement is definitely warranted – this technology is likely to cause a step change in the advance of knowledge, and its application, for good or ill.

2.2 The excitement is generated, in part, by the nature of the machine changes going on and, in part, by the great speed of those changes.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

3.1 **It is recommended** that the Select Committee should note, that it is often the scientific and technical leaders in AI who raise concerns about the effects that AI is already having, and may yet have.

3.2 **It is recommended** that the Select Committee consider if the very rapid advance of AI capabilities calls for a modern version of the Asilomar Conference 1975. A non-expert’s background information is found here https://en.wikipedia.org/wiki/Asilomar_Conference_on_Recombinant_DNA. By this is meant, that when the technology of recombinant DNA was invented and found to be extremely far-reaching, the scientists, technologists and commercial interests involved persuaded each other to stop further work until they, and

wider society, had had a chance to consider ethical, safety, legal and other societal issues. Perhaps previous experience can be a guide now.

3.3 **It is recommended** that the Select Committee enquire, of technical experts, what steps might already be underway as modern analogies of the “Asilomar” process.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

4.1 People with money to invest are most likely to gain and those with low wealth will suffer severely. Perhaps those who are made redundant should be given a significant number of shares in the company they are leaving in a way that they are unable to sell them for a long period but can benefit from the dividends. Perhaps all could be given purpose built unit trusts that will deliver dividends. Displaced employees could thus, more readily, choose whether they work or not or, perhaps, take up charity work or other means of taking up their time.

4.2 Work is satisfying and essential to the wellbeing of most people at the present time. We need to redefine how we usefully and satisfactorily use our time. If work, in the historical sense, becomes impossible for all due to lack of jobs then what can we do to fill our time, which is accepted and endorsed by society? At the moment society expresses the view that work is good and unemployment is bad. Government has worked to reinforce that view. This is not an imperative, society could recognise that there is not enough historical work to go round and recognise also that we are not really defined by our job.

4.3 **It is recommended** that the Select Committee consider how the role of Government could help in the process of redefinition of work.

4.4 **It is recommended** that the Select Committee consider if this is the time to undertake trials of a “Citizen’s Income”. In which case, the Select Committee should consult the Parliaments of Finland, Switzerland and Holland, in whose countries such trials are either underway or have been considered.

4.5 If the Select Committee’s remit does not embrace recommendations 4.3 and 4.4 **It is strongly recommended** that the Select Committee refers these points to an appropriate alternative Parliamentary forum.

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

5.1 The experience of the Raymond Williams Foundation (RWF), more widely known in the North of England than in the South, is useful in this context. This adult education charity aspires to the aim of the late Raymond Williams: “I’ve often defined my own social purpose as the creation of an educated and participating democracy”. Our community-based discussion circles have a sound

track record of meetings, courses and workshops on themes which many voters would normally think are “highbrow” and dominated by elites. Against this, RWF fosters a local, devolved, grass-roots, participative democratic culture, open and non-sectarian. The huge implications of the AI debate have already been tackled in this network. These community-based debates will continue, not only in our evening and day discussion circles and groups, but also within more ambitious day and residential events, when major players, authors and speakers may be invited to stimulate and lead the debates.

5.2 It is apparent, from discussions around our educational network, that the more distant a person is from the subjects of science, technology, engineering and mathematics (especially statistics) – the so-called STEM subjects – the less likely s/he is to appreciate the changes that are underway. There is a lesser tendency for some people, distant from STEM subjects, to dismiss news of AI as exaggeration or as science fiction and so to disengage from this area of public concern.

5.3 Education is crucial, of course with schools, colleges and universities and adult education, which should all be encouraged to engage with AI, using Select Committee guidelines and information resources. Adult education and, especially now, the wide and growing informal networks such as Philosophy in Pubs (PiPs); Discussion in Pubs (DiPs); Café Philosophique; faith group circles, etc (see RWF website <http://www.raymondwilliamsfoundation.org.uk> for list and detail on all these) will continue to have AI on the agenda. The discussion tools and guidance on all this are freely available on the RWF website. They could easily and cheaply be extended for wider public promotion, supported by government at all levels, but without compromising the freedom of individuals and each group ‘to follow the argument wherever it leads’ in a non-party and non-sectarian fashion. This is possibly a good model for 21st century adult education, on all big issues.

5.4 **It is recommended** that the Select Committee, and through you the Government, engages vigorously with the world of informal adult education, as an essential part of improving the public understanding of, and engagement with, artificial intelligence.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

6.1 We are not competent to express a view on this question.

7. How can the data-based monopolies of some large corporations, and the ‘winner-takes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

7.1 This question produced a sharp disagreement within our network of respondents, as expressed in 7.2 and 7.3, below.

7.2 Some, in our network of respondents, argued that data ownership and privacy are not really associated with the impact of AI, and should not have been included in the Select Committee's consultation paper. Privacy and data ownership are already an issue, without AI, and should have been dealt with separately. Their inclusion could cloud the real issues associated with what is essentially a new intelligence arising to compete with humans. If this intelligence had arisen biologically and evolutionally, conflict would almost certainly ensue. Conflict will not arise with AI only if we as humans perceive that we are in total control of this new entity. If the perception is that we are being superseded then, somewhere at some time, a group of opposing militants will arise.

7.3 Others in our network of respondents argued that data ownership and privacy are closely associated with the impact of AI. It was considered that the tendency of commercial digital enterprises to lay claim to data ownership is the modern parallel of the "enclosure" of common land some two centuries ago. Indeed the phrase "data enclosure" has been coined. It is so central to AI, that enterprises based on the manipulation of vast quantities of personal data appear to be valued on stock markets by the quantity of data they "own" rather than by reference to physical metrics such as turnover or profit.

7.4 **It is recommended** that the Select Committee should examine how citizens might be enabled to exercise the right of access to any data held about them by commercial or governmental organisations. Such an examination may well have an international dimension and policy change may need international agreements.

Ethics

8. *What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?*

8.1 There are already more than 9 million robotic machines operating in the world. Commonly known examples are machines that build other machines, AI warehouse transport vehicles and autonomous road vehicles. There is a present and pressing need for public discussion of, and legislation to address the results of accidental damage and injury. For example, a recent death on an assembly line worker is cited here: <https://qz.com/931304/a-robot-is-blamed-in-death-of-a-maintenance-technician-at-ventra-ionia-main-in-michigan/>.

8.2 **It is recommended** that the Select Committee consider who is legally to blame following an accident involving an autonomous machine. Is it the machine itself, the person nominally in charge, the owner, designer, manufacturer, and supplier or a third party?

8.3 An autonomous machine may have the capacity to decide what to do in the event of unforeseen circumstances. It is important, even now, that society should reassert that all human lives and wellbeing are equal and autonomous machines must be required to act on that principle.

8.4 **It is strongly recommended** that the Select Committee reasserts that, in all interactions between autonomous machines and humans, all human lives and wellbeing are equal.

9. *In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?*

9.1 We are not competent to express a view on this question.

The role of the Government

10. *What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?*

10.1 The government should monitor the progress and effects of AI very closely with a well-resourced agency. It should try to anticipate trends, both good and bad. Plans can be made, but what will be really important will be reacting properly to those trends that have not been anticipated. AI and its effects will be very complex and society can't hope to foresee everything that will occur. It is important, therefore, that hitherto unknown trends are picked up early and reacted to quickly. The governmental tendency to react after the horse has bolted will certainly not do in this case.

10.2 A specific example of regulation is in the potential for the formation of emotional relationships between humans and AI machines. It is easy enough for animal-human emotional bonds to form. In Japan, it is already apparent that humans can form emotional bonds with autonomous machines.

10.3 **It is recommended** that the Select Committee should consider British Standard BS 8611:2016 (*Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems*). Although it does not refer specifically to the formation of machine-human emotional relationships, the use of British Standards would potentially form part of a range of public tools to regulate AI.

10.4 A second specific example is the rapidly increasing sophistication of machine-generated speech, machine-generated translation and machine speech recognition. Once these technologies merge, if not before, it will be essential that humans will need to be told, by some means, if the voice they are interacting with is human or machine. This will be especially important for those with impaired hearing: currently about 20% of UK pensioners are fitted with hearing aids and another 20% have seriously impaired hearing.

10.5 **It is recommended** that the Select Committee should consider if new law is required to ensure that humans are told when they are interacting with machine-generated speech.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

11.1 Japanese society appears to be more at ease with human-like manifestations of AI than many other societies. The proportion of human-like robots to the population, in Japan, is way beyond what has happened so far in the UK. **It is recommended** that the Select Committee should seek advice from Japan, including from its parliament.

11.2 Both Iceland and Estonia have moved far faster than the UK in adopting advanced data handling systems as integral tools to build democracy. **It is recommended** that the Select Committee should seek advice from Iceland and Estonia, including their Parliaments, in considering the potential impacts of AI and big data handling on democratic processes.

6 September 2017

Professor Chris Reed – Written evidence (AIC0055)

Evidence to the House of Lords Select Committee on Artificial Intelligence

Chris Reed, Professor of Electronic Commerce Law, Centre for Commercial Law Studies, School of Law Queen Mary University of London

1. This response addresses questions 8-10 of the Call for Evidence, focusing on transparency and thus particularly on question 9. It examines how a lack of transparency impacts on existing legal mechanisms, and how (and in what ways) the law might either change to accommodate uses of artificial intelligence (AI), or make demands about the use of AI. It focuses particularly on the use of machine learning techniques in the production of AI systems.
2. The research on which this evidence is based was undertaken by the Microsoft Cloud Computing Research Centre and is available as Queen Mary University of London, School of Law Legal Studies Research Paper No. 243/2016.¹¹⁰⁹

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

3. Commenting on the ethical implications of artificial intelligence is largely outside my domain of expertise as a lawyer, though there is clearly a multitude of ethical implications. It is, perhaps, worth noting that the law tends to subordinate itself to society's collective view about what kinds of behaviour are, and are not, ethical. This makes the ethical framework particularly important.
4. As an example, take the liability of a doctor for making an incorrect diagnosis. Whether it is ethical for a doctor to substitute the diagnosis of an AI system for his or her own human judgement is a complex question. The question arises even if the AI is demonstrated to produce more accurate diagnoses, on average, than a human doctor, because the AI may produce incorrect results in cases where a human doctor would have delivered a correct diagnosis. However, once the medical profession has reached a consensus on this point, liability law will largely accept that consensus. Thus if a substantial body of doctors believe that this kind of reliance on an AI is acceptable practice, the law will hold that the doctor acted reasonably and is thus not liable in negligence.
5. Sometimes the ethical consensus of society is embedded in the law. The obvious example is that of fundamental rights, such as the rights to free speech, privacy, and non-discrimination. The law thus contains a normative statement about how humans ought to behave, but of course many do not behave in that way, otherwise there would be no need for the law. This mismatch between the law's

¹¹⁰⁹ Chris Reed, Elizabeth Kennedy & Sara Nogueira Silva, 'Responsibility, Autonomy and Accountability: legal liability for machine learning' Queen Mary University of London, School of Law Legal Studies Research Paper No. 243/2016, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2853462.

ethical demands and actual behaviour becomes problematic when an AI incorporates machine learning. The decisions made by the AI will reflect the actuality of human behaviour rather than meeting the normative demands of the law. This mixed legal and ethical issue will be addressed further in response to question 9.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

6. Technical systems whose workings are not understandable by humans are often described as 'black box' systems. It is of fundamental importance that the law recognises that a system might be a 'black box' to one person but not to another. For example, the producer of a machine learning AI might be able to explain how and why it reaches its decisions, whilst to the user of the technology these matters will be unknowable.
7. This distinction is important because the application of the law often depends on what a human knew, or ought to have known, at the time the liability arose:
 - In most cases, all that the user of the technology knows is that he is ignorant of its workings, and that it is *de facto* a 'black box'. His only options are to rely on its decisions or reject them.
 - A producer of AI technology is in a different position, however:
 - The law will ask what a technology producer knew or ought to have known in advance, for example through the process of testing and evaluating the technology.
 - It will also ask what can be discovered after the event if the technology fails to make a correct decision.
8. Thus any legal requirement to incorporate transparency into an AI needs to take account of these differences in perspective. Merely demanding transparency is meaningless, because the law regulates human activity (and, in this context, human decision-making in particular). It is therefore essential to define a requirement for transparency in terms of the human who needs to understand the decision-making processes of the AI in question.
9. Many elements of law rely on transparency because they require persons whose decisions have caused loss or damage to have acted reasonably or fairly. If there is evidence that the outcome of a person's decision was unreasonable or unfair, it will be necessary to justify those decisions or face liability. An obvious example is the liability of a car user, where the car incorporates autonomous driving technology. In the current state of the law the user will be liable if it was negligent to use that technology in the circumstances, or to operate it in that particular way (eg through being unable to retake control of the vehicle in circumstances where the technology hands back control). In theory, the human car user needs sufficient transparency about how the AI makes its decisions so that, in turn, the human can decide whether and how to use the technology. In practice, humans tend to become very reliant on apparently working technology, so continuing to base the liability of car users on their (presumed) state of knowledge may be inappropriate.

10. Transparency works differently in the case of an AI producer. Here the question will be whether it was negligent for the producer to put the technology on the market. The law's inquiry will focus on three areas of knowledge:
 - How far the data on which the AI was trained is an accurate representation of the situations which are likely to confront the AI in use.
 - The extensiveness of the testing regime for the AI, and in particular how closely the testing related to the risks which are foreseeably likely to arise from using the AI.
 - Any known sets of circumstances in which the AI performs less well than the human decision maker it is intended to replace, and how well the producer has taken steps to deal with such underperformance when the AI is in use.
11. On the assumption that any AI which is put into use will have been demonstrated to perform better, on average, than a human decision maker, a legal liability enquiry will be likely to focus primarily on the training and testing regimes. The existing law on liability incentivises producers to keep detailed records on these matters and also to preserve the training dataset, and these will be vital in meeting any transparency obligations.
12. Currently the law does not require producers of technology to explain how that technology works. A motor vehicle manufacturer could, for example, incorporate an innovative braking system without any obligation to explain its workings to the car user. Imposing a requirement on those who produce or supply AI technology to provide these kinds of transparency would therefore be legally novel.
13. In the case of an AI, the most useful form of transparency is being able to explain the basis of and reasoning behind the AI's decisions. There is an important distinction to be made between ex ante transparency, where the decision-making process can be explained in advance of the AI being used, and ex post transparency, where the decision making process is not known in advance but can be discovered by testing the AI's performance in the same circumstances. Any law mandating transparency needs to make it clear which kind of transparency is required.
14. In the absence of a wide range of real-life examples of AI in widespread use it is difficult to identify any fundamental principles which could be used to determine whether a transparency obligation should be imposed, but some tentative starting points for such a discussion are proposed here:
 - 14.1. **Complete lack of any transparency**, ex ante or ex post, as is currently the case for some AIs based on neural networks. This should be legally acceptable where the AI produces benefits to society overall and the loss to individuals is minor and compensatable. An example might be an AI controlling a domestic central heating system, where the only risk to the householder is that use of the AI results in higher bills than before it was installed. In cases like this the law's policy decision might reasonably be to maintain the current legal position, which is that the householder bears this risk, sharing it with the supplier and producer of the technology via the existing law on liability for defective products and services. The

situation seems no different from the introduction any other new, but non-AI, technology.

It is conceivable that new uses of non-transparent AIs might create risks of loss which is not easily compensatable under the current law, but it is too early to predict what form these might take, and whether some special liability scheme might need to be devised to cope.

- 14.2. **Lack of ex ante transparency.** This should be legally acceptable where the AI produces benefits to society overall and loss to individuals is legally compensatable (ie monetary damages will be accepted by society as an adequate remedy). This is effectively the current position for personal injury caused by motor accidents. Society accepts that some injuries are inevitable, and that human decision-making when driving is unpredictable, and also accepts that monetary compensation is the best that can be achieved while still permitting the societal benefits from motor travel.

However, although such a lack of ex ante transparency is acceptable as a matter of legal principle, there are likely to be AI implementations where it is unacceptable to society at large. Autonomous vehicles are one example – when I have presented on this topic the audience has generally expressed reluctance to run the risk of death or injury without some prior justification (usually to a regulator) of the reasoning through which the AI decides between potential victims in an inevitable crash. A show of hands gives strong support for the proposition “I’d rather be killed by a human than a machine”, even though those voting recognise this is irrational if the AI reduces the overall risk of injury. Law embodies the social settlement, not merely abstract principle, and so may need to impose requirements for ex ante transparency even if they are not, from a purely legal perspective, necessary.

Against this, though, any ex ante transparency regulation needs to recognise that the regulation will reduce the ability of an AI to improve in use via machine learning. An AI which is required to provide ex ante transparency cannot evolve its decision-making through learning, but instead will need to capture use data and upload that to the producer’s training set. The AI can then be trained on this data, and in due course an improved version released, but this is inevitably slower than evolution through learning in use. Ex ante transparency regulation might also prevent the use of non-algorithmic AIs such as those incorporating neural networks, because it is usually not possible to explain ex ante (if at all) the reasoning through which the neural network reached its decision.

- 14.3. **Lack of ex post transparency** is likely to be unacceptable except in those cases (see point 1) where the potential loss is minor and easily compensatable. Ex post transparency is essential to improve the future performance of an AI and to reduce the risk of reoccurrence. Only the

producers of an AI technology are likely to be able to provide ex post transparency, and thus any transparency regulation should focus on their systems of training and testing, recordkeeping and data retention. Ex post transparency is needed if the existing law of negligence is to be used as the mechanism for deciding liability, because without such transparency there would be no way of deciding whether the accident resulted from a lack of care by any human. In practice, though, the cost and difficulty of obtaining this evidence (particularly if the AI producer is not a UK company) will make using the existing law more difficult and expensive. However, the law of negligence is capable of evolving to deal with this problem.¹¹¹⁰ More simply, though, a strict liability scheme could be introduced backed by compulsory insurance, as proposed by section 2 of the Vehicle Technology and Aviation Bill 2017 (and see the research paper in note 1, though written before the Bill's publication, for more detailed discussion of this point).

14.4. Ex ante transparency is needed where fundamental rights are at risk. Most countries have a range of laws which protect fundamental rights, such those which prohibit discrimination on the grounds of race, sex, sexual orientation, religion, etc. These laws do not apply to machine learning technologies directly, but rather to those persons who are using the technology to assist them in carrying out another activity. The problem is that an AI based on machine learning will embed the decision-making which infringes those rights in real life, rather than the ideal behaviour which is described in the law. The potential for infringement of the right to a fair trial by using an AI to assist in sentencing has been recognised in *State of Wisconsin v Loomis*¹¹¹¹, and the practice of motor insurers granting insurance policies to women on more favourable terms than to men, because the statistical evidence is clear that women present a lower risk, has been held unlawful on the ground of sex discrimination by the Court of Justice of the European Union.¹¹¹²

The only way to ensure that such infringement of fundamental rights is not implicitly occurring when an AI makes decisions is to require the reasoning leading to those decisions to be explained ex ante. How to impose such a requirement is a difficult question though, because for many AIs it is not

¹¹¹⁰ Most likely by creating a presumption that some person (probably the AI producer) was negligent unless that person could prove that reasonable care had been taken.

¹¹¹¹ (2016) WI 68.

¹¹¹² *Association Belge des Consommateurs Test-Achats ASBL v Conseil des Ministres* (Case C-236/09, 1 March 2011). Employment decisions, such as shortlisting for interview, are increasingly assisted by machine learning and would give rise to liability on the same basis – see eg Chen-Fu Chien & Li-Fei Chien, 'Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry' (2008) 34 *Expert Systems with Applications* 280.

obvious that they risk making infringing decisions until an actual infringement occurs. This is explored further in the answer to Question 10.

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

15. First, it is inappropriate and impossible to attempt to produce a regulatory regime which applies to *all* AIs. The range of potential applications is far too diverse – it is obviously foolish to apply the same regulatory regime to autonomous vehicles and also to smart refrigerators which order groceries based on consumption patterns. Indeed, there is no plausible, let alone compelling, reason to regulate smart refrigerators at all. Second, without some years of real-world experience of the use of AI technologies it is debatable what transparency might mean when translated into law and regulation, let alone how it could be achieved. The history of legislating prospectively for the digital technologies is one of almost complete failure.¹¹¹³
16. This suggests that the Government should adopt a risk-based approach, regulating only those uses of AI which present both immediate and high risks, until it becomes clear how the technology will evolve and be used. It also raises the question how far the Government should attempt directly to regulate the production and use of AI, or whether it should instead create incentives to introduce transparency indirectly, by means of changes to existing legal regimes such as the legal liability to compensate for losses.
17. The strict liability regime when a motor vehicle is 'driving itself', proposed by section 2 of the Vehicle Technology and Aviation Bill 2017, is a good example of how the latter approach might incentivise transparency. Under section 2(1) the insurer of the vehicle is liable for damage caused by the vehicle in an accident, irrespective of any negligence on the part of the vehicle user. But section 5 allows the insurer to claim against any other person liable for the accident, which in this case would include the vehicle manufacturer or the producer of the AI which undertook the driving. To defend against such a claim, both manufacturer and AI producer would need to be able to provide ex post transparency to explain how the accident occurred, or risk a court finding that the accident most likely was caused by their negligence. Thus section 5 clearly incentivises ex post transparency.
18. However, indirect incentives might be insufficient to deal with society's fears about some new AI technologies, with autonomous vehicles being the obvious example. If direct regulation of AIs is necessary for this reason, the regulatory focus should be on the proper construction of learning data (including its preservation) and the testing of the AI's decision-making quality (including keeping full and appropriate records). This focuses on the elements of an AI's development where problems are created, and provides the information necessary to attempt to ameliorate those problems. Any proposed regulation requiring ex ante transparency should consider carefully the potential effects

¹¹¹³ See Chris Reed, 'How to Make Bad Law: Lessons from Cyberspace' (2010) 73 MLR 903.

which it might have in limiting the use of particular AI technologies, and also on the ability to improve the decision-making quality of the AI.

19. As previously explained, there is a strong theoretical argument for requiring ex ante transparency if an AI's decisions have the potential to infringe fundamental rights. In practice, identifying in advance those AI implementations which have such a potential is extremely difficult, so that any regulation outside those sectors of activity which are already known to present fundamental rights risks is likely to be either over- or under-inclusive.
20. The safest approach, if the beneficial effects of AI implementation are not to be discouraged, is to proceed initially by way of the existing legal regime. This allows affected individuals to seek a remedy against public authorities for any fundamental rights infringement resulting from an authority's decisions¹¹¹⁴, and for specific rights claims may be made against private sector persons.¹¹¹⁵ The possibility of successful claims will incentivise potential defendants to ensure that their use of AIs does not lead to infringements, and they will thus demand from AI producers sufficient transparency to reassure them on this matter.
21. However, the effect of AI use on fundamental rights needs to be kept under review. There is a distinct possibility that some AI implementations might result in a large number of minor fundamental rights infringements where the damage to individuals is not sufficiently great to make a legal claim worth pursuing. There would thus be no incentive to seek transparency or to amend the AI so that it no longer made infringing decisions. If fundamental rights truly *are* fundamental, it must be impermissible for known infringements to be allowed to continue merely because the technology which produces the infringements has other (primarily financial and operational) benefits.

Professor Chris Reed

3 September 2017

¹¹¹⁴ Human Rights Act 1998 s. 8.

¹¹¹⁵ For example, claims for sex discrimination in employment under Equality Act 2010, ss 64-80, 120-126.

Research Councils UK – Written evidence (AIC0142)

1. **Research Councils UK (RCUK)** is the strategic partnership of the UK's seven Research Councils¹¹¹⁶. Our collective ambition is to ensure the UK remains the best place in the world to do research, innovate and grow business for the benefit of society and the economy. Together, we invest more than £3 billion in research each year, covering all disciplines and sectors. This response is made on behalf of the seven Research Councils¹¹¹⁷ and represents their independent views.

Summary of key points

- AI systems currently have narrow functionality. Generalised AI is still a long way off.
- AI could affect jobs, creating opportunities, but providing support for continuing education and re-skinning will be important in realising these opportunities.
- Sectors that could see benefit include: transport, finance, insurance, retail, legal, healthcare, manufacturing, environment, agriculture.
- There are societal and ethical issues (including the need for responsible research and innovation) that need continuing research.

¹¹¹⁶ www.rcuk.ac.uk

¹¹¹⁷ www.ahrc.ac.uk; www.bbsrc.ac.uk; www.epsrc.ac.uk; www.esrc.ac.uk; www.mrc.ac.uk; www.nerc.ac.uk; www.stfc.ac.uk

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Artificial Intelligence is the theory and development of computer systems to perform tasks normally requiring human intelligence. AI technologies seek to reproduce or surpass abilities that would require 'intelligence' to perform including learning and adaptation; reasoning and planning; sensory understanding and interaction; optimising parameters and procedures; autonomy; creativity; and knowledge extraction and prediction from large, diverse digital data.

Contemporary AI is less about making machines think like humans and more about engineering approaches to develop machines that engage in the rational behaviour associated with humans achieved by means which are quite unlike human reasoning.

The importance of AI has grown in recent years due to: continuing increases in information and data available for AI systems to analyse and learn from; advances in algorithms; the increasing capability of computing hardware; and the ubiquity of mobile high-power computing.

Globally, AI adoption is occurring faster in more digitised sectors and across the value chain. This is most apparent in high technology, telecommunications and financial services sectors. However, AI is being deployed in many other areas, although it is often unevenly adopted across assets, labour and usage components. For example, the use of AI support for digital assets associated with Consumer Packaged Goods and Transportation and Logistics sectors. Other areas include autonomous vehicles, smart robotics, virtual agents, natural language and computer vision, healthcare, agriculture, decision making software, and online advertising and recommendations. However, the use of AI for machine learning¹¹¹⁸, for both multi-use and non-specific applications, accounts for the single largest area of investment. (McKinsey Global Institute: *Artificial Intelligence – the next digital frontier?* June 2017).

The next 5 years could see deployment in industry and commerce of AI-enabled decision support systems (for example see World Economic Forum *Shaping the Future of Production* March 2017).

¹¹¹⁸ Machine Learning is an approach to AI that explores the study and construction of algorithms that learn from and make predictions on data. The growth in available data has driven major advances in machine learning that have seen it broadly adopted.

The next 10 years will see continuing adoption of AI systems across a broad range of computer systems. AI will increasingly be embodied in Robotics and autonomous systems that work alongside people. This will be especially true within constrained scenarios (e.g. in agriculture, mining or warehousing). Use of autonomous vehicles could have quickest uptake in commercial services (e.g. logistics, taxis), but is likely for most large automotive manufacturers in all areas.

The next 20 years, deeper support for higher reasoning is possible and the productivity of most workers could be enhanced by AI assistants.

There are limitations, however, as a 'general' AI system has yet to be developed. AI technologies currently have narrow functionality. Systems designed for one application cannot be applied to another and the performance of systems is influenced and limited by the quality and availability of training data. AI is also limited to a basic understanding of human emotions. The concepts of 'social intelligence' and the ability to 'compute-with-meaning' are goals of human-like AI, wherein an AI is able to comprehend and interpret context, thus proceeding to adapt accordingly in its responses.

Future challenges include the need for both hardware and software to adapt to handle the increasingly vast amounts of data available for information and knowledge management systems to store, and for AI systems to process and utilise. This may lead to a rise in hardware-friendly AI, or development of AI technologies that are friendly to hardware. A major challenge is to develop technologies that are more energy efficient than currently available, whilst retaining their overall efficacy.

Other technical factors include:

- Current reasoning systems are complex and expensive to set up, requiring specialist skills, careful data preparation, and time.
- AI systems are less strong when data are limited, uncertain, and inaccurate, leading to inflexible applications and sometimes unsafe results when confronted with new situations.
- Current machine learning approaches are poor at explanation generation; providing a reason for a decision can be hard.

We also need to explore the economic and social implications of our increasing use of robotics, automation and AI. These are likely to emerge over a fairly short period of time. There are still many unanswered questions around issues of work, inequality and ethics that have the potential to impact dramatically on the development and diffusion of AI.

Automation requires both the algorithms and training in their use. Decision-making algorithms (for example insurance, healthcare, criminal justice) are not free of the innate biases of human decision-makers and can make unfair or discriminatory determinations. They this have the potential to create new types of harms. Furthermore, datasets often reflect latent social and technological biases, thus bias and discrimination may merge only when an algorithm processes particular data. This can further embed and amplify discriminatory and unjust structures in society rather than challenging them. We must be able understand how a conclusion might be reached over the fairness, bias and transparency of an algorithm ie how will these concepts be defined in this context where what is fair for one may be unfair for another.

There are many unanswered questions around issues of work, inequality and ethics that have the potential to impact dramatically on the development and diffusion of AI:

- **Work:** AI has the potential to create new industries and employment opportunities, but there is also a risk. Estimates suggest that about 15 million current jobs in Britain are at risk from automation over the next couple of decades. Although technological revolutions are not new jobs and wages could change faster and more fundamentally than in the past.
- **Inequality:** Unmanaged proliferation of AI technology could exacerbate inequalities between different segments of the UK population across a wide range of areas including access to economic opportunities, education/training, healthcare and political participation. Moreover, the resources and expertise required for the Big Data approach to AI is likely to concentrate economic power in the hands of a relatively small number of organisations and companies.
- **Ethics:** There are concerns around data privacy, ownership and protection. Decision-making delegated to machines also raises issues of responsibility and oversight.

It is critical that technological development proceeds with an understanding of its wider social and economic implications and the capability and willingness to address its effects.

2. Is the current level of excitement which surrounds artificial intelligence warranted?
--

The current level of excitement is warranted. Indeed, the UK Government has recognised that AI is important and offers gains in efficiency and performance to most industry sectors. It has commissioned a major AI review led by Wendy Hall and Jerome Pesenti to identify the critical elements for the technology to grow in the UK and consider how government and industry could work together.

However, adoption could be slower than sometimes promoted because of the complexity of establishing applications. Also the capability delivered is less than sometimes promised.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

Employment: Areas which have experienced relative growth in recent decades are likely to be the most affected. This includes jobs focussed on driving vehicles, call centres, warehousing and distribution, clerical work and routine decision making across the service industries. Longer term, managerial and decision making posts are also likely to be affected but the extent is unpredictable.

Skills: There is a shortage in the skills required to build, maintain and to assess and operate AI systems: mathematics, algorithms and software development, and the domain skills to develop and apply the algorithms. Managerial and oversight structures may currently be inadequate to handle these new systems. The AI Review notes that STEM graduates will have the fundamental skills, but a wider range will be needed in the AI workforce as it increasingly overlaps with ethics and social sciences. For example economists are needed for the development of fintech systems, linguists for the development of language processing systems. Jobs will change in response to pervasive use of the latest technology. This could mean that in-career re-skilling will become the norm every 10 years.

Education: There is a need to educate the public on the impact, opportunities and limitations of AI. It is important we start to prepare for the skills needed. This starts in schools and school education needs to prepare future generations to be more adaptable.

Privacy: Given the need for data to be made available for AI systems, there is a need to inform people of how their personal data are being consumed. This may require revisiting data privacy and data protection legislation to allow for repositories that may operate on personal data with the intention of learning general models, along with strong assurance that the data shall not be used for any other purpose. A significant barrier to the development of AI systems is concern over the privacy and protection of sensitive data. In healthcare, finance, and many sectors (in particular those serving individuals directly) data need to be protected for reasons of privacy, security, confidentiality and commercial sensitivity. Data-holders need assurance and trust to be confident in sharing data

with AI developers. Some of the areas where data are most sensitive may also be ones where the greatest benefits lie.

Cyber-security: AI has enabled machines to learn the typical conditions to be seen on a company network, and spot anomalous behaviour. It has the advantage of detecting unknown threats without needing to know the nature of the threat, and can act to mitigate the potential impact. There is also greater potential for the incoming threats to be powered by AI systems. Investment will be needed to counter this.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

AI could benefit anyone at any level of society. Hall and Pesenti believe impacts will be positive, large and spread across sectors. AI will also augment individuals (deepening our memory, speeding recognition) and generally (by reducing mundane tasks) allowing better decision making. However this needs investment. Creating an AI is expensive, requiring large computing and data storage resources and a skilled workforce so large companies are better placed to take advantage which concentrates expertise into large corporations. With this economic dominance power can be exerted to control and exploit markets, and unduly influence opinion. Regulation is required.

Moreover, AI is becoming more and more opaque and with new “languages” being invented that the majority of people struggle to understand. Algorithms are also becoming increasingly complicated, so it is difficult to understand from the code whether they have inherent biases. Some systems rely on an element of randomness and produce nondeterministic outputs. This has profound ramifications for everyone across all aspects of our lives and for democracy more widely. Social science is particularly important helping to understand the social constructs that underpin the data and AI and algorithms.

Creativity, social intelligence and the ability to interact with complex objects in unstructured environments are human traits that are difficult for machines to learn. Jobs which require these qualities are likely to be less susceptible to wholesale replacement by AI systems than those that contain routine cognitive work or physical labour.

Between the extremes of narrow and general AI there is “specialised AI”. Technologies here include machine learning enabling knowledge-based companies to make a step change in their productivity, by augmenting the skills and expertise of humans letting each do what they are best at and using new

processes to create more customer value. This could lead to disruptive business models.

Development and use of artificial intelligence and data is regional. Localities with lower levels of investment in technological and digital infrastructure and low skill levels are likely to be hardest hit by AI technologies. Investment is needed to access the rewards of adoption of AI.

The rapid pace and adoption of AI makes the identification of emerging disparities difficult to predict. Research is required to understand the source, nature and potential consequences of these disparities and this should be undertaken as a broader approach to AI innovation.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

AI can pose a societal challenge as it can be perceived as a disruptive technology allowing the replacement of the human element by automating tasks and may (according to Stephen Hawking) pose a threat to humankind's survival. However, AI should be viewed as a technology that shares the same world as humans, behaving as a 'hybrid agent' augmenting human endeavours by promoting independence and enhancing productivity. There is a need for public dialogue on the impact, opportunities and limitations of AI, so it is seen as not truly intelligence, but a complex decision maker.

Issues that need to be addressed include:

- A lack of trust in the reliability and fairness of computer mediated decisions
- The need for human mediation and appeals procedures
- Providing explanation generation to increase trust
- Clarifying the legal framework in which AI operates
- Responsible representation in the press.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

Sectors that will benefit are those where the bulk of the work is decision making of a repetitive sort. Areas of potential benefit include:

- Transport and other areas which require use of vehicles
- Finance
- Insurance
- Retail
- Logistics
- Legal
- Healthcare
- Manufacturing
- Environmental modelling and prediction to inform decision-making across sectors
- Agriculture
- Extreme and hazardous environments such as energy (nuclear, off-shore), deep mining and space.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

The monopolies of Uber and Facebook are due to network effects; that of Amazon to investors who tolerate one but not many fast-growth low-profit companies. Data are valuable and UK supermarkets, for example, will not share their data from store cards with competitors. As larger companies have smaller unit costs, a 'winner takes all' economy has a variety of root causes, not just data monopoly.

Actions that can be taken to mitigate the risks include:

- Using education and training to develop and spread skills
- Investing in AI to develop independent expertise and capacity such as science and research, infrastructure modelling, government policy.
- Clarity and openness in the rights and responsibilities of data controllers.

- Strengthening data protection and transparency; providing clarity on the rights of data subjects on the use of data about them.
- Noting that publicly funded research data are a public good and should be made available with as few restrictions as possible; and that there may be legal, ethical and commercial constraints on their release.
- Maintaining network neutrality to ensure a level playing-field.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

There are ethical implications to the development and use of AI. Algorithmic decision-making has the potential to harm individuals in new and unique ways where existing ethical and legal frameworks struggle to provide solutions. For example, when individuals are discriminated against not because of their relevant characteristics, but because they match a specific type identified by an algorithm (for example, residents of low-income areas). The increasing complexity of algorithms and the lack of appropriate AI skills could mean that we build black boxes that make decisions based on biased data which have serious implications for individuals and society.

AI also raises important ethical questions around:

- Data collection, privacy, consent and ownership: to whom does the data belong? Can consent be withdrawn and records raised? What are the rights of individuals?
- Construction of algorithms: who decides if the decision algorithm is ethical, who checks this?
- How are data used and for what purposes? Can individuals control how data about them is used, what for and by whom?
- Who is ultimately responsible for algorithmic outputs? The developer, the data analyst or the operator?
- Is there transparency around how data are collected and used? How are AI and its outputs regulated? How will misuse be monitored and punished?

The Research Councils are conscious of these issues and encourage researchers to incorporate responsible research and innovation in the programmes¹¹¹⁹.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable? When should it not be permissible?

Decision-making algorithms need regulation, verification and validation to understand their efficacy and to reduce bias. The lack of explanation can undermine trust in agencies leading to resistance to the uptake of decisions. Systems need to be transparent so citizens know the criteria being used.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

The Government has a role in:

- Promotion of skills and capacity by investing in activities and resources which can develop and expand AI capability, expertise and capacity within the UK. Examples under direct government control where investment could be directed include: science and research, infrastructure modelling and development, government policy and agencies.
- Development of ethical regulatory frameworks clarifying responsibilities and liabilities with respect to AI mediated decision making e.g. in healthcare and finance.
- Political and electoral funding and regulation, e.g. balance and accuracy.
- Privacy, data ownership and transparency of data use. These will need a responsive and proportionate regulatory framework which builds on current good practice.
- Promotion of open data as a public good where it does not contradict privacy and data protection concerns.

AI systems are subject to regulatory measures in, for example, data protection, health and safety, legal. AI techniques need to be consistent with existing regulations and these, in turn should reflect AI approaches.

¹¹¹⁹ Responsible Innovation is a process that seeks to promote creativity and opportunities for science and innovation that are socially desirable and undertaken in the public interest.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

- The EU values on data protection, data ownership and privacy. For example, the General Data Protection Regulation provides a basis for use of data relating to individuals including safeguards to ensure personal data is used appropriately and remains secure. These should be adopted and continued after leaving the EU.
- The EU also places an emphasis on competition and reducing monopoly power in the technology sector.

The UK should align itself with EU and WEF policy: these strike a balance between public protection and permitting scientific research.

Research Councils UK

6 September 2017

Research into Employment, Empowerment and Futures Centre (REEF), The Open University – Written evidence (AIC0124)

Summary

S1. This response has been prepared by Peter Bloom, Evangelia Baralou, Vincenza Priola, Owain Smolovic-Jones and Pinelopi Troullinou – all members of REEF. It draws on, but is not confined to, emerging recent research into the cultural, organisational, and political implications of artificial intelligence (AI) on society.

S2. This wide-ranging research focuses on diverse possibilities for AI to empower and potentially disempower individuals and communities across contexts throughout the UK and internationally. In particular it draws upon multidisciplinary methods to study themes of technology and democracy, inclusion, social mobility and the potential for broad based economic transformation. In this response:

- We show that the scope of short, medium and long term change associated with AI is profound – potentially affecting all populations though in uneven ways that have the capacity to either promote equality or deepen existing inequality (Questions 1, 2, 4, 5, 7, 8, 9 and 10).
- From our analyses we reveal the need for greater attention to the ideological and political dimensions of these AI driven shifts for the promotion of social inclusion, democratisation, professional development, personal wellbeing and empowerment both individually and at the community level (Questions 2, 3, 5, and 10).
- We find that there is a lack of public knowledge and tools for realising the full extent of this transformation – especially in how it can enhance the public good and significantly alter the current ‘winner takes all’ economic system (Questions 2, 7, 10, and 11).
- We suggest that there is presently a gap in the public use of AI to make a positive social impact locally, regionally and nationally and we highlight how it could lead to further dramatic positive changes in the future (Questions 3, 4, and 5).

About REEF

R1. REEF was established at The Open University (OU) in April 2017, and draws on the centre’s multi-disciplinary scholarly expertise on themes of identity, leadership, power, human relations, social and work inclusion and learning to allow policy makers, organisational leaders, social practitioners and people from

across contexts to work together to co-create innovative solutions for enhancing the empowering economic, political and social effects of emerging technologies.

R2. REEF is based in the OU's Department for People and Organisations. Research within the department seeks to explore and shape the future direction of empowerment, work and society. It aims to help individuals and communities to develop the skills and knowledge necessary to engage positively with a precarious present and future.

R3. REEF draws on expertise across the university, including that of the Knowledge Media Institute – the lead partner in the recently completed MK:Smart¹¹²⁰ project – which has won several awards including three at the Smart Cities UK 2017 awards (in Data, Communications and Energy).

R4. With nearly 174,000 students, the OU is the largest UK University. It operates across all four UK nations and through academic research, pedagogic innovation and collaborative partnerships, seeks to be a world leader in the design, content and delivery of supported open learning.

Question 1 What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

1.1 The 21st century is on the verge of a possible total economic and political revolution. Technological advances in robotics, computing and digital communications have the potential to completely transform how people live and work. Even more radically, humans will soon be interacting with AI as a normal and essential part of their daily existence. Crucial now more than ever is the importance of rethinking social relations to meet the challenges of this soon-to-arrive 'smart' world.

1.2 Research being conducted by members of REEF (Bloom and Clarke, forthcoming, 2018) proposes an original theory of trans-human relations for this coming future. This theory will be characterised by a fresh emphasis on infusing programming with values of social justice, protecting the rights and views of all forms of 'consciousness' and creating the structures and practices necessary for encouraging a culture of 'mutual intelligent design'. It involves moving beyond the anthropocentric worldview of today and expanding current assumptions about the state of tomorrow's politics, institutions, laws and even everyday existence.

¹¹²⁰ <http://www.mksmart.org/>

1.3 The next 5, 10 and 20 years will specifically see the following developments in these areas:

- *Meaningful Intelligence*: Human and artificial intelligence will continue to combine to produce deeper forms of meaningful intelligence. AI and robotics will be further developed to enhance personal growth and wellbeing. These efforts may become even more acutely important as new technologies enable humans to live and machines to operate for longer. This makes it even more critical to discover how human and non-human relations can assist each other so that humans lead not just longer but also fuller more compassionate and caring lives.
- *Creating Smart Economies*: There is growing potential for humans and machines to use their shared meaningful intelligence to build 'smarter' and more egalitarian economies locally and globally. AI and automation is meant to make economies and societies 'smarter' – more efficient, productive and convenient. Yet smart technology also holds the promise of ushering in a 'post-work economy' where the need for labour is reduced and material scarcity is a thing of the past. However, for these utopian visions to be made into a reality requires the use of non-human capabilities and intelligence to create an economy that is liberating and just as it is smart. In fact the risk is the marginalization of the poorer class, which may remain cut off from the direct benefits AI can bring.
- *Mutual Intelligent Design*: The rise of AI has spurred new political visions of technological emancipation spanning the ideological spectrum from libertarianism to socialism. Yet both its detractors and proponents miss the full radical political potential of a trans-human politics. It is one where non-humans and human citizens deploy the latest virtual, digital and manufacturing advances to mutually design their societies.

1.4 These potentially exciting short, medium and longer term developments are being hindered by widespread and legitimate trepidation over fears that AI will be a 'disruptive' force creating mass unemployment and social dehumanization. The current so-called digital divide is threatening to become an even bigger 'future divide' where those with access to and training with AI will benefit from it while those without such advantages will experience its negative effects. These concerns are currently limiting the potential positive social applications and impact of AI.

Question 2 Is the current level of excitement which surrounds artificial intelligence warranted?

2.1 The excitement surrounding AI is certainly warranted. However, it must also be treated with caution equally in terms of what is driving this public enthusiasm, what is producing less positive responses and finally what is being critically ignored. These concerns point to broader issues of whether the full social

potential of AI is being publicly discussed and exploited and whether its benefits are being widely shared or are primarily limited to technologically and economically privileged populations.

2.2 The public appetite for AI is clearly growing. Smart technology is creating new products seemingly by the month that improve the speed of communication and the ease with which we can buy products and organize our lives. Yet this excitement is tempered by fears that AI will be used to invade our privacy and replace us economically. It is further limited by the relative lack of awareness about how fundamentally AI can progressively transform society in line with desires for greater welfare, democracy and equality.

2.3 Current research by REEF members (Troullinou, 2017; Bloom, 2013, 2016) reveals that presently, the public responses to AI can generally be categorised as either 'seduced' or 'resistant' – neither of which lead to particularly empowering results. In the first case, people are psychologically and materially drawn to new technologies choosing to ignore or minimise its threats to their wellbeing and broader social concerns. In the second, they wholly reject these advances thus losing out from experiencing their potential positive benefits.

2.4 Fundamental to overcoming these unproductive responses is to stress that humans have the power, individually and collectively, to shape the future of the AI. The current discourse of globalization as 'inevitable', for instance, has led to widespread public backlash, and an apparent resurgence in nationalism and negativity about multilateralism and international cooperation in the context of Brexit. For this interest to be maintained and broadened there is a pronounced need by policy makers to ensure that this change is both empowering and transformational.

Question 3 How can the general public best be prepared for more widespread use of artificial intelligence?

3.1 As discussed under Question 1, recent technological developments in AI, robotics, digital communication and distributed manufacturing are transforming the world. At the level of everyday practice, there is increasing reliance on AI to assist and complete tasks such as navigation, driving, scheduling, shopping, consuming, job seeking and even lifestyle preferences. Through its enhanced data analysis, AI is helping to guide everything from where you should go on holiday, to what you should eat, how you should exercise and what career opportunities you should pursue. Economically, it is already altering design, manufacturing and distribution of goods and services. For example, applying AI on new techniques like 3D printing are making manufacturing and selling more localised, globally connected, operationally flexible and personalised to users' needs.

3.2 However, there is an important need to combine these advances in smart technology with public policies that promote values of social inclusion and community empowerment. A current proposed REEF project (Bloom and Baralou, 2017) seeks to realise this goal through creation of 'Future Labs' – collective spaces for individuals and communities to directly use technology to explore the exciting ways these innovations can profoundly alter and improve our current society and economy. This project would aim to make technological change empowering by drawing upon serious gaming and virtual reality to show people the full extent of its social possibilities for creating a world that is not only 'smarter' but also more equitable and just.

3.3 Specifically, we would create a simulated 'future world' that would allow diverse communities to experience a 'day in the life' of different AI enabled futures. We would also produce 'future games' that would permit people to take part in public ownership of municipal resources, hi-tech co-operatives and participatory 'smart' community budgeting. Finally, this project would work with disadvantaged communities and marginalized populations to invent virtual worlds so that policy makers, service providers and the wider public could experience potential future everyday realities thus highlighting specific social, economic and political barriers to success, providing those in power with greater knowledge to more effectively tackle these issues while also fostering wider social unity amongst diverse populations.

Question 4 Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

4.1 Existing research shows that currently the primary beneficiaries of AI are large corporations, technology based businesses, governments and affluent consumers. Questions remain regarding the effects of AI on labour and the already widening gap between the richer and the poorer. A member of REEF, Peter Bloom, was the lead academic on a recent BBC Documentary *Secrets of Silicon Valley* revealing how AI driven companies like UBER and Air BnB profit from weakening labour rights and government regulations. Additionally, employers have disproportionately benefited from AI enabled techniques that enhance their ability to collect and use employee data to increase workforce efficiency, sometimes at the expense of the psychological and physical wellbeing of their employees.

4.2 The people gaining the least from these developments are those with little or no access to this digital technology and no opportunities to use it for their own personal advancement or community improvement. This includes those in economically disadvantaged areas and regions as well as the poorer social class and groups such as ethnic minorities, women, those with a disability and those in the LGTB communities who face systematic personal and professional prejudice.

Recent evidence uncovers widespread class, racial and gender bias in the tech industry and within AI programming (as REEF member Cinzia Priola reports in *Inequalities, discrimination and the pay gap. Can HR play a role in fostering workplace inclusion and equality?*, upcoming blog for HR most influential). Quoting REEF member Owain Smolovic-Jones in a recent opinion piece for industry platform People Space: “The future of work is a choice but business needs to step up and be brave many parts of the UK have been in a post-work situation for generations now. Whole generations of people in the 1980s and onwards experienced their jobs becoming obsolete due to privatisation and globalisation...automation has the potential to unleash similar effects – on the same people but also new waves of people”. These findings highlight the growing ‘future divide’ discussed in Question 1 – one that limits the benefits of these technological advancements to diverse populations along with the public good that they can deliver to society as a whole.

4.3 There are number of key ways to mitigate and ultimately bridge this emerging ‘future divide’. An upcoming REEF produced feature length television documentary explores these solutions. A potentially strong idea with increasingly popular support is implementation of a universal basic income to reduce the negative impact of AI driven economic ‘disruptions’ – particularly linked to unemployment. There is also the need to create wider programmes of direct involvement of technology users in development of the technologies themselves (e.g. of disabled people in development of prosthetics, wheelchairs, learning tools, etc.). Education programmes from an early age, could include direct exposure to programming and robotics. Also, redirecting lifelong learning so individuals can engage in continual processes of ‘future skilling’ is paramount. Further benefit would come from public education campaigns to explore changes to the present economic system away from a ‘winner takes all’ market model to an empowering post-work scenario. Lastly, these benefits would be more widely distributed through technologically enhanced practices and policies of workplace democracy and employee rights that would provide for a wider voice on the direction of travel regarding uses of AI.

Question 5 Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

5.1 The use of AI has wide public and individual benefits based on the new capabilities emerging from this technology’s internet based interconnection and real time generated data. However, these benefits must be weighed against a range of public concerns such as those linked to data privacy. It is important to strike the right balance between empowerment and protection to maximise positive engagement with AI by individuals and communities. There is increasing research evidence that public understanding and engagement with AI is not universal. Different demographics have different levels of knowledge and

different concerns based on their particular needs, cultural values and existing relationship to technology. For this reason, it is crucial that data empowerment projects are socially inclusive and sensitive to these diverse empowerment discourses and privacy concerns.

5.2 In order to ensure that AI is empowering and respectful of people's privacy members of REEF (Bloom, Troullinou and Priola, 2017) are developing a community based model for balancing values of data privacy and data empowerment (potentially as part of the grant programme supported by the Information Commissioner's Office). The project will develop the model by direct engagement with diverse communities and populations to understand their particular data privacy concerns and data empowerment needs. It will integrate innovative distance learning methods with face to face interactions to allow three often marginalized groups (young women, religious minorities, and impoverished communities) to collectively create their own custom data streams and apps.

5.3 This project will have significant benefits for increasing public understanding and engagement with the empowering potential of AI. Notably, it will create a community based approach for promoting data empowerment and addressing data privacy issues. It will, furthermore, provide traditionally marginalised social groups with the opportunity to use AI in a 'hands on' way, having a positive impact on their communities while enhancing their personal knowledge and skills associated with AI. More widely, REEF members will create a general course and toolkit for diverse communities to learn how to take advantage of the empowering possibilities of AI. Finally, this project will raise the need for a 'bottom-up' perspective for the broader public engagement with AI – linked to issues such as data privacy and data empowerment – within the broader public debate. This is just one example of how community programmes could be developed to engage groups of individuals who are generally socially marginalised and could be further marginalised by further developments in AI and robotics.

Question 6 What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

6.1 AI has the potential to benefit all sectors including health, domestic life and education. However the key sectors that stand to report the largest economic benefits from its development and use are (as discussed in Question 4) private multinational corporations, those in the technology industry and those organisations across sectors that benefit from state surveillance policies. They have the capital, access, and expertise to develop and best exploit technologies. However, their focus primarily rests on producing, marketing and selling more commercial products as well as finding more sophisticated ways to gather information on employees and citizens. Underdeveloped, in this respect, are the broader positive contributions AI can have for local economies, public service and

broader issues of democracy, social inclusion and civic engagement. The need to extend the benefits of AI to the overall public good rests with governments and public institutions and with potential partnership between these and the technology industry.

6.2 There are a number of sectors that are not presently benefiting from AI to their full capabilities. These include individual entrepreneurs (particularly those from disadvantaged communities and populations), small and medium sized businesses (SMEs), and local governments across the UK. Their barriers include financial resources, access, technological expertise and knowledge. Research being conducted by REEF – as evidenced in our upcoming documentary (see Question 4) – reveals that this lack of benefit is particularly ironic in light of the fact that AI in the areas of manufacturing, product design, and problem solving positively encourages ‘global’ business strategies and personalised ‘batch based’ production that would fit well into the organisational models characteristic of these sectors and stakeholders. Issues remained focused on questions of access and involvement of different social groups.

6.3 In order to improve access and use of AI by these sectors, REEF is presently in discussion with corporations, foundations and local councils and regional administrations to assist SMEs and disadvantaged people bridge the ‘future divide’ through networking with existing ‘future entrepreneurs’ and technologists to discover how to use and develop AI based techniques such as open access technology, digital fabrication and distributed manufacturing for their wider short and long term benefit. We are also in the initial stages of collaborating with projects like MK:Smart that promote the innovative public use of AI for improving local public service and solving chronic community based problems linked to congestion, health access and energy use.

Question 7 How can the data-based monopolies of some large corporations, and the ‘winner takes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

7.1 There is a crucial need to address current data based monopolies of some large corporations and the broader ‘winner takes all’ economies associated with them. It is absolutely critical to expand the possibilities for empowering ‘post-employment/post-work’ communities in the present through a range of projects and education programmes. This would help vulnerable communities and populations to take an active role in innovative solutions to problems of chronic unemployment and social exclusion. It would also offer leading stakeholders the tools and skills necessary to promote and implement a future empowerment agenda in their organisations and communities.

7.2 Toward these ends, REEF is leading on creation of a multidisciplinary and wide ranging empowerment agenda for the future. In particular, this focuses on ways to expand the meaning of empowerment to include ways to produce a progressive 'smart' society and a 'post-employment/post-work' economy that is inclusive and whose benefits are universally accessible to all. REEF's free online course, *Modern Empowerment at Work*, explores for a wider audience how emerging empowerment perspectives and strategies are redefining contemporary organisations and management. The final week of this four week course explicitly examines future challenges and opportunities for empowerment linked to technological change.

7.3 REEF is also in the process of setting up a 'sustainable futures' consortium composed of private, public and not-for-profit organisations across the UK that are interested in developing policies, practices and knowledge around how to use AI and other emergent future technologies for fostering an empowering present and future. The goal would be to create models that can be applied across a wide range contexts for 'bridging the future divide' and moving away from a technology driven 'winner takes all' economy. These organisations would not only share knowledge but, with REEF's steer, publish policy suggestions and offer nationwide workshops to spread these ideas more broadly.

Question 8 What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

8.1 The ethical implications of the development and use of AI are increasing rapidly. They include current issues of data privacy, human disempowerment, AI's negative economic impacts and future ethical concerns such as those focusing on the existential nature of AI powered machines and robots and their overall treatment. In particular, there are growing concerns that people have little say in shaping either their present or future. These worries are heightened by predictions that technology will soon dramatically 'disrupt' how we currently live and work. Already, these fears can be seen in the insecurity experienced by precarious workers and the alienation many feel in relation to digital technology. While there is much talk about the need to prepare for these 'disruptions', there is little discussion about what values, structures and practices can be fostered to help people do so in a constructive and engaged way. There is also a need to explore the implications of these developments for the theory and practice of democracy within organisations as well as within civil society.

8.2 These negative implications can be resolved through committing to an interconnected three pronged approach developed by REEF:

- *Creating Open Futures Together*: How can we ensure that the benefits of the future will be accessible to all and inclusive of everyone's voice and needs? What are the broader opportunities for people to explore

collaboratively existing and new possibilities that would enhance their overall wellbeing?

- *Engaging with a Precarious Present*: How can we build institutions and relations that allow people to approach change with excitement rather than anxiety?
- *Learning and Producing Knowledge Today for an Empowering Tomorrow*: What are the present and future skills people need to feel empowered within a precarious present and an uncertain future? What types of innovative and technologically advanced knowledge and teaching approaches can we develop to help people take advantage of these existing and future opportunities and challenges?

Question 10 What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

10.1 The government has a large role to play in maximising the overall social benefit of AI. Research shows that AI is viewed as a potential social and economic 'disruption'. This discourse of disruption plays into a broader contemporary 'crises narrative' in which present uncertainty is coped with through trying to 'recover' the perceived stability of the past (Bloom, 2015). This paradoxical narrative was initially associated with the 2008 global financial crash but most recently has perhaps fuelled an apparent resurgence of political nationalism (e.g. 'Make America Great Again'). The same narrative is now also influencing the reception of AI linked to an unpredictable and threatening future. Governments have a responsibility at all levels to promote an empowering vision of AI alongside concrete local, regional, national and even global policies aimed at realizing this empowering vision. There is a significant potential for governments to influence the directions of AI development and use, not necessarily in terms of the technologies themselves, rather in terms of supporting engagement with them and broadening their use. Above we have included a few examples of programmes that the UK Government could directly or indirectly support, however within a wider agenda for social inclusion many more could be developed.

10.2 Early research conducted by members of REEF reveals the need to transform AI 'disruptions' from a threat to an opportunity. This means changing the narrative into a positive vision of coming social and political change. In terms of AI, this requires crafting a viable public story highlighting the role of this technology for contributing to social, economic, and political progress through positively 'disrupting' a problematic status quo.

6 September 2017

Professor Kathleen Richardson and Ms Nika Mahnič – Written evidence (AIC0200)

Authors:

Professor Kathleen Richardson
Centre for Computing and Social Responsibility, De Montfort University, Leicester
Founder of the Campaign Against Sex Robots

Ms Nika Mahnič
Masters Student in Big Data
King's College London
Writer for the Campaign Against Sex Robots

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

AI has been invigorated by digital culture but its reinvigoration as a practice has not come from any breakthroughs in replicating human intelligence in machines, but in creating more manipulative marketing tools for big companies and corporations. This has meant that the push to life online has provided a new means for companies to sell products to consumers.

AI is a tool, just as robots are tools and they are designed to help humanity to have a better life. Those who propose that AI is humanlike drawn on a degraded and partial view of human beings that is rooted in the idea that peoples' bodies and their relationships are no different from inanimate things and relationships between things. This perspective allows researchers in AI to view humans as forms of property capable of commercial exploitation (slavery). This degraded view of human beings is then reproduced in narratives of robots and AI.

Artificial Intelligence should be renamed Advertising Intelligence because is a largely using a commercial model of capitalism as the template for the human beings in co-relationship with each other.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Companies that have the financial resources and experts for data mining and the development of algorithms used for enhancing businesses benefit the most. Moreover, such powerful companies are often institutionally attached to universities and research agencies with enough money, specifically the ones funded by industry. As it was and is with oil (data, the fuel for AI, is said to be the new oil), it is the ones who have access to tools for extraction and manipulation of »natural« resources (in the case of data, human behaviour on social media) that have the opportunities to benefit from data manipulation and analysis. Data is the prerequisite for artificial intelligence working well and being beneficial for companies – the general population`s complicity with using free services is in many cases a consequence of insufficient knowledge of the processes fuelling digital monopolies, such as Facebook and Google and their business models based on data.

It is of utmost importance to make people aware of the nature of the “free services” they use daily, what they are giving away with giving away their privacy, their behaviour in exchange for free digital services, how this enables profits of companies having monopolies in the digital economy, as well as how this might impact their future life opportunities. Using Facebook and Google is a Faustian pact: the right to be forgotten remains in the realm of wishful thinking. Digital traces are being used against citizens, predominately for marketing and propaganda. Looking at the fringes of society, to the ones not worthy of the citizen subjectivity – refugees - we understand that the tools used against them in the name of security (IBM) are only a beginning of using data for an even more advanced social sorting of general population. Refugees as the test subjects in the new algorithmic regulation regime that is limiting mobility of people in the times when mobility of goods, capital (and data) stays uninhibited.

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

Unfortunately research scientists win prestigious EU funding and other source of income by telling myths about the potential of AI (and robots) that are in fact false. Huge EU funded projects are now promoting unfounded mythologies about the capabilities of AI. When these projects fail or do not deliver the results, the researchers ask for more resources and often get the resources. This means that our resources get ploughed into mythical research avenues based on hyperbole rather than real human good.

We believe that the premises at work in every AI project should be thoroughly investigated. If the premises are anti-human, racist, sexist, or ageist, they should not be supported for funding. Only a human-centred approach that values human beings can move the science of AI into a beneficial direction.

Finally, EU projects now support many ethics components. Most of these philosophers are white middle class men, who often engaged in sexist ethics against women. Also, these ethics researchers are not encouraged to actually challenge the science or technology. If ethics of the human is to be a valuable contribution to the way we develop technology, ethics cannot be reproduced by a small subsection of the population who benefit from the status quo.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

The simplest way to put it would be – development and use of artificial intelligence benefits the profits of companies in every sector, but is detrimental to informed, not-manipulated political activity. The latter would be the specifically non-industrial sector, for example politics perceived as a service (not a sector) made to benefit citizens. This argument could be countered by saying that Facebook and Google serve the citizens by developing “civil tech”, such as voting reminders or thematic search during campaign time (Facebook) or sorting useful information and exposing it (Google – that is at the same time hiding extremist – anti-neoliberal content).

Some propagate beneficial effects of automated CV sorting, automated grading, credit risk grading and other phenomena that use artificial intelligence for social sorting, or in the case of Facebook and Google, social engineering powered by algorithmic manipulation. When optimisation and profitmaking are the main reasons for using technology for such purposes, we need to ask how much does a computerized error cost – in the case of abovementioned “sectors”, it costs human lives and automation does not being enough profits to be justified.

7. How can the data-based monopolies of some large corporations, and the ‘winner- takes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

If we name the main actor, the practices of curating reality and history by Facebook should be monitored – for example, the company`s cleaning (in the name of monitoring) people`s Newsfeeds of problematic information should be public. For example, the rules of what presents problematic data should be publicly available and non-compliance with their own rules should be publicly problematised, and wrongdoings punished by censorship. Every monopoly, and this should also count for digital ones, should be monitored at international courts – moreover, Terms and conditions should be presented in a way that enables *informed* consent.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

We root our ethics in the politics of anti-slavery. Only in resistance to slavery has a version of the human developed that is not a tool and should not be treated like an object.

Therefore we believe all ethics should be rooted in narratives of anti-slavery, so that sexism, racism and classism are not reproduced in new ethical models of AI. Unfortunately this is what is now happening across Europe. Researchers in AI are wondering if machines are slavery or robots should have rights. This completely disrespects the history of humanity who have suffered and toiled for others, often in oppressive regimes. We oppose this move to break down the distinctions between humans and machines. This is an urgent priority.

7 September 2017

Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy and Professor Simon King – Written evidence (AIC0029)

**Professor David Robertson, Dr Michael Rovatsos,
Professor Austin Tate, Professor Chris Williams,
Professor Robert Fisher, Professor Alan Bundy and
Professor Simon King – Written evidence (AIC0029)**

[Submission to be found under Professor Robert Fisher](#)

Mrs Violet Rook – Written evidence (AIC0151)

QUESTIONS.....

The pace of technological change and the development of artificial intelligence

The impact of artificial intelligence on society

The public perception of artificial intelligence

The sectors most, and least likely, to benefit from artificial intelligence

The data-based monopolies of some large corporations

The ethical implications of artificial intelligence

The role of the Government and The work of other countries or international organisations.

The development of artificial intelligence indicates the progress of society and can help with the growing needs of education, health and industry via the collection of data and the analysis of information. It can promote positive change in regard to all these topics, but there are aspects which need scrutiny if in the future artificial intelligence is to help not hinder progress.

Yet AI can also aid the promotion of gambling and ideas which prove harmful to society. The data-based monopolies of some large corporations leads to individuals being stereotyped and categorised for the sake of easy production and profit making models. Algorithms are made by human beings and therefore artificial intelligence can be made to give advantage to some while being biased to others for example in determination of jobs and insurance. Many organisations use computers in this manner to make decisions about clients which can be based on variables which seem to be balanced.

I herewith list below the Positive and Negative aspects involved within society of AI.

Artificial Intelligence-positive aspects

Information is easier to obtain for education and health purposes.

The world becomes a smaller place and encourages travel.

Communication is promoted via mobile phones and computers therefore of benefit to individuals, government and industry.

Artificial Intelligence-negative aspects.

Algorithms are made by humans therefore could have an in built bias in regard to gender, age, race, religion and other protected characteristics. Without scrutiny provided by law this could distort results and prove dangerous to the individual and society.

The selling of data and making this a funding stream which seems to be legal puts privacy and security of the individual in doubt despite Data Protection Laws.

Summary: The topic is vast and needs review on a regular basis if the public is to feel confident in the methods used in regard to Big Data and the use of artificial intelligence to provide a civil society. It is often clear that sometimes the term Artificial Intelligence is used to indicate robotics with the uses of data and the

Mrs Violet Rook – Written evidence (AIC0151)

methods of collection via artificial intelligence and algorithms in everyday life being overlooked.

6 September 2017

Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy, Professor Simon King and Professor David Robertson – Written evidence (AIC0029)

Dr Michael Rovatsos, Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy, Professor Simon King and Professor David Robertson – Written evidence (AIC0029)

[Submission to be found under Professor Robert Fisher](#)

Royal Academy of Engineering – Written evidence (AIC0140)

Summary

1. There are many successful applications of modern artificial intelligence (AI), both in physical applications such as robotics and autonomous systems, and in non-physical applications. While current applications use 'narrow' AI that is focussed towards very specific applications and tasks, 'general' AI represents a much greater scientific and engineering challenge, although there have been some advances. In the future, the use of AI may help to tackle some of the structural and infrastructure challenges facing society. All sectors of the economy stand to benefit from artificial intelligence.
2. The acceleration of the state of the art over the past 10 years has been profound, and has resulted in a high level of excitement that could lead to both dystopian and utopian perspectives on what is likely to happen. Although neither of these extremes is likely to occur, continued open debate, insightful thinking and careful responses are required. Both the private sector and governments will have to move quickly to create sensible governance and control.
3. At this stage of the development of AI, it is mainly pioneering organisations that are active and experimenting. This is generating evidence of how best to create value that the fast-followers will start to exploit, at which point the use of AI will grow further, with corresponding economic benefits.
4. New types of data will continue to emerge as physical systems such as infrastructure are increasingly controlled by software and information about human activity becomes available. This will lead to new applications of AI. Organisations will need to develop the ability to make best use of their data, as well identifying opportunities to share data with other organisations.
5. As huge quantities of data are held by a few companies, there is a risk that the benefits of AI will accrue to a limited number of players if mechanisms are not put in place to encourage competition. Government needs to recognise that big data is a common good and the fuel of the AI transformation.

6. AI will impact society by displacing some jobs, enhancing human engagement in others and creating new employment and leisure time opportunities. Skills is a key issue and action needs to be taken now because of the length of the education pipeline. High-level skills are required as well as skills for people who can understand the potential of the technology in business or other areas. For the broader population, technical and data literacy should be taken as seriously as the 3Rs.
7. There are some commentators that present worst-case scenarios that capture the public's imagination and present challenges to the wider uptake of AI. A concerted public awareness campaign on both the benefits of AI and actions being taken to mitigate the downsides is needed, emphasising how AI in partnership with humans can be more productive.
8. Government, businesses and public bodies will need to consider their use of AI in decision-making, consulting widely, and ensuring that mechanisms are in place to detect and address any mistakes, biases or unintended consequences of decisions made. Ensuring transparency of algorithmic decision-making is a challenge, particularly for machine learning and self-adaptive systems.
9. While the regulatory landscape is developing, government should lead by example by applying standards to its own use of AI, to ensure accountability and help build public trust in use of algorithms. It is important to ensure that regulatory guidance and criteria are developed with sufficient expert input. The Academy stands ready to advise government on regulatory issues, as appropriate.

10. Introduction

11. The Royal Academy of Engineering welcomes the opportunity to respond to this call for evidence on artificial intelligence. As the UK's national academy for engineering, the Academy brings together the most successful and talented engineers from across the engineering sectors for a shared purpose: to advance and promote excellence in engineering. The Academy's response has been informed by the expertise of its Fellowship, which represents the nation's best engineering researchers, innovators, entrepreneurs, and business and industry leaders.

The pace of technological change

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

12. There are many successful applications of modern artificial intelligence (AI), both in physical applications such as robotics and autonomous systems, and in non-physical applications. Its use is growing in a wide range of sectors such as smart cities and intelligent mobility, advanced manufacturing, energy, finance, law, entertainment, education and healthcare. Additionally, companies providing AI services are emerging¹¹²¹. Novel applications are being developed where new forms of data, often in vast quantities, are being brought together with AI techniques¹¹²². Advanced data analytics will increasingly exploit AI technologies since the volume of data generated will require approaches that can learn patterns and highlight results autonomously – a key role for AI. Many new applications will emerge that have not yet been envisaged with the potential to provide novel solutions to many of society's major challenges. While there are pockets of innovative practice developing, the challenge will be to spread examples of early best practice.
13. The acceleration of the state of the art over the past 10 years has been profound. This is important because it means both private sector and governments will have to move quickly to create sensible governance and control. Furthermore, this governance will need to be global in nature, rather than national, which makes implementation even more difficult.
14. A number of factors have converged to make this pace of change possible, including breakthroughs in AI techniques such as neural networks, greater access to computers with significant processing power, the generation and ability to access large data sets and a greater level of investment in the technology.
15. New types of data will continue to emerge, as well as innovations in the way different datasets are used in combination. This will accelerate as associated technologies such as the Internet of Things, cloud services,

¹¹²¹ For example, Google Deepmind are developing artificial intelligence techniques and their applications, especially in healthcare www.deepmind.com; Vivacity are applying machine learning to smart cities and intelligent mobility www.vivacitylabs.com; Feedzai are using banks' data on customer behaviour to apply predictive techniques to assess risk in real-time. Machine learning is used to detect anomalies and thus very subtle signs of fraud www.feedzai.com; CloudNC is using AI methods in manufacturing to control CNC milling machines automatically <https://angel.co/cloudnc>; Energi Mine uses artificial intelligence software to procure and trade energy www.energimine.com/.

¹¹²² The Planet and Orbital Insight partnership combines Planet's satellite datasets (broad coverage, high frequency monitoring from nano satellite constellations) with Orbital's automated geo-analysis for business <https://www.planet.com/pulse/planet-strikes-landmark-deal-with-orbital-insight-to-address-financial-markets/>

personal mobile devices and data platforms develop. Correspondingly, new applications for AI will emerge that previously were not possible without this data, and without access to information about human activity or infrastructure that can be controlled by software. Goods and services that use AI in their operation are likely to be gathering their own data and combining it with data received from elsewhere. They may also share all or part of their data or results with other parties, triggering opportunities for new goods and services and related economic benefits.

16. While government has led the way with open data initiatives, much potentially valuable data remains locked away in corporate silos or within sectors¹¹²³. Much of the issue is that this data is not curated or understood within most organisations, thus they are not able to make good use of the data they have, let alone make it available to others. The Academy is currently engaged in a study examining the opportunities and barriers for data sharing and trading based on case study research and would be happy to share early findings. Furthermore, access to large data sets is currently dominated by a small number of companies, with the risk that the benefits of AI will be held in a limited number of players.
17. Potential applications will continue to present themselves over the next 5 to 10 years, although the rate at which this happens will vary according to the specific context. For example, constraints on access to data and how it is used may be very different for healthcare applications making use of electronic patient records, say, compared to business or industry applications using corporate data – each has its own challenges. There may be varying levels of acceptability and adoption across applications which are influenced by a combination of social, economic and cultural factors. Where AI is used to control physical systems – for example, in autonomous vehicles – there are specific ethical and legal issues that could affect adoption, such as the potential shift in responsibility for safe operation from operator to designer. Cybersecurity is another challenge; breaches in security could result in a loss of integrity to algorithms as well as the data they make use of, with resulting safety and privacy implications¹¹²⁴.
18. The range of AI techniques is broad¹¹²⁵. Current applications use 'narrow' AI that is focussed towards very specific applications and tasks. 'General' AI represents a much greater scientific and engineering challenge,

¹¹²³ Royal Academy of Engineering and IET (2015), Connecting data: driving productivity and innovation, <http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity>

¹¹²⁴ The Academy is currently engaged in an ongoing programme of work on the cyber safety and resilience of critical infrastructure and the internet of things.

¹¹²⁵ Artificial intelligence can be subdivided into a number of categories. Some is deterministic and its outputs predictable even if very complex. Other artificial intelligence systems are non-deterministic, learning systems and these can be further split into those such as neural networks which are frozen before release to users and those which continue to learn after release.

although there have been some advances¹¹²⁶. Active cross-disciplinary collaboration between, for example, computer scientists and neuroscientists is helping to push the state of the art, and narrow AI is beginning to apply lessons learned from one environment to another. However, one of the central challenges in achieving general AI is 'transfer learning' – the ability of computers to infer what might work in a given scenario based on knowledge gained in an apparently unrelated scenario - which is not something they currently can do. Although the timescale for general AI goes well beyond the next 20 years, it is critical to understand now how to solve the 'control problem'¹¹²⁷ as it is such an important issue.

19. Solving key research challenges is a high priority so that machine learning and other AI systems can be deployed safely, securely and effectively. These include ensuring outputs can be interpreted and the workings are transparent, creating systems whose behaviour can be predicted and verified with high confidence, building systems that discover causal relationships and not just correlations, and ensuring systems are not vulnerable to cyberattack¹¹²⁸.
20. At a national level, the use of AI may help to tackle some of the structural and infrastructure challenges resulting from, for example, climate change or population growth. It will allow us to model, understand, anticipate and respond to adverse weather conditions, changing water levels, energy needs and use, terrorist plots, waste management and human safety in specific situations. All of these uses will increase our security, improve the use of resources and create stability for economic prosperity. A focus on 'benefits-led' innovation, combined with better data sharing and connectivity, will help cross-sectoral innovation to occur, resulting in the economic, social and environmental benefits that AI has the potential to bring.

Is the current level of excitement which surrounds artificial intelligence warranted?

21. Alan Turing argued¹¹²⁹ that there is no reason why a computer cannot perform all the functions of a human brain. Nothing has emerged since 1950 that fatally undermines Turing's argument¹¹³⁰. So the current interest

¹¹²⁶ SingularityHub (March 2017), *Google chases general intelligence with new AI that has a memory*, <https://singularityhub.com/2017/03/29/google-chases-general-intelligence-with-new-ai-that-has-a-memory/>

¹¹²⁷ Nick Bostrom (July 2014), *Superintelligence: Paths, dangers and strategies*.

¹¹²⁸ Royal Society (April 2017), *Machine learning: the power and promise of computers that learn by example*, <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>. Chapter 6 describes the key unsolved research challenges.

¹¹²⁹ in his seminal 1950 paper (Computing Machinery and Intelligence, *Mind*, 59 433-460)

¹¹³⁰ Stephen Hawking and others wrote in May 2014 in *The Independent* that "there are no fundamental limits to what can be achieved: there is no physical law precluding particles from

in AI is certainly warranted, although perhaps based on unrealistic expectations about the speed of progress.

22. Compared to previous rushes of enthusiasm for AI, this time round it is founded on practical results rather than wishful thinking. For example, the Alvey programme¹¹³¹ in the 1980s was driven by the fear of a Japanese fifth generation computer threat that never materialised¹¹³². There is compelling evidence that current AI techniques based on access to large scale computing and information resources are applicable to a wide range of applications, in contrast to earlier approaches that delivered results that could not be transferred from one application to another. To a significant degree this is because the approach is based on automated learning rather than human-driven programming.
23. The smartphone is another technology development that is enabling the wider use of AI. This device is packed with sensors and provides a very convenient way to collect data and feed it back to cloud-based AI applications. The fact that a phone can see, hear and locate itself makes it a very powerful device. Moreover, since a large proportion of the population carries a smartphone, AI can demand our attention extremely easily.
24. As with most technological advances, the impact in the short term will be lower than anticipated, but in the long term every device used by people, and the controls attached to infrastructure, tools, vehicles and buildings, will include some notion of adaptive and learning behaviour.
25. The level of excitement is high enough to lead to both dystopian and utopian perspectives on what is likely to happen but neither of these extremes is likely to occur. However, the existence of these perspectives reflects the profound, and very fast-moving, change that is occurring that will require continued open debate, insightful thinking and careful responses.

Impact on society

How can the general public best be prepared for more widespread use of artificial intelligence?

26. Artificial intelligence is already used routinely in everyday life, for example in intelligent personal assistants that use AI voice recognition such as Amazon's Alexa, Apple's Siri and Google's Home, and by Netflix and others for movie recommendations. It is not always visible and that will continue

being organised in ways that perform even more advanced computations than the arrangements of particles in human brains”.

¹¹³¹ This was a British government sponsored research program in information technology that ran from 1983 to 1987, and that focused on artificial intelligence among other things.

¹¹³² NY Times (June 1992), *'Fifth Generation' became Japan's lost generation*

<http://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html>

to be the case for a number of AI applications. If done well, most consumers will not be aware that it is AI, but only that the things they use work more effectively and consume less resource. Other applications will be more visible.

27. AI will impact society by displacing some jobs, enhancing human engagement in others and creating new employment and leisure time opportunities, although there are varying views on the impact of these changes. The Industrial Digitalisation Review¹¹³³ is addressing the impact of digitalisation on industry. While there could be a displacement of jobs in the short term, industrial digitalisation has the potential to create new, better-paid jobs – both in developing suitable AI techniques and other emerging technologies, and in operating and maintaining them in particular contexts. Furthermore, the growth that would result from increased productivity and innovation could potentially lead to the creation of additional jobs. In many cases, such technologies will be used to enhance the role of humans rather than replace them. The pace of change is such that dislocations could be very painful and conventional solutions may not adequately address the problem. Some have warned against complacency by professionals such as doctors, lawyers and accountants, whose jobs could be replaced by less-expert people, new types of experts and high-performing systems¹¹³⁴.
28. Skills is a key issue and action needs to be taken now because of the length of the education pipeline. Many of the high-level skills required in AI – for example, for interpreting data or avoiding bias – are common with data science, and there is a shortage of such skills. There is also a skills gap for people who can work with an AI system but are not AI experts. These people understand the potential of the technology and its limitations and can see how it might be used in business, but are not in a position to advance the state of the art.
29. It will be important to include ethics in any training. The challenge of ensuring diversity and inclusion in the workforce and of reducing digital exclusion will need to be addressed. Those people whose jobs are displaced should have the opportunity to retrain. Furthermore, a much broader social and political rethink will be required that considers what 'good work' means for people, and how active and positive citizenship is recognised and rewarded in an environment where most young people currently entering the workforce will not have a long settled career.

¹¹³³ Industrial Digitalisation Review - Interim Report (July 2017), Version 3.0

http://industrialdigitalisation.org.uk/wp-content/uploads/2017/07/Interim_Report_Final3_1.pdf

¹¹³⁴ Richard and Daniel Susskind (October 2016), *Technology will replace many doctors, lawyers, and other professionals*, Harvard Business Review, <https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals>

30. While a detailed explanation of how an AI algorithm made a decision is beyond the understanding of non-experts, greater awareness of how the technology around us works is needed, and this should be addressed in schools, and in accessible ways - such as television, courses and websites - for adults. Technical and data literacy should be taken as seriously as the 3Rs since a broad population with these skills will be necessary to support industry and other economic activity.
31. The Academy is identifying the challenges of digital skills in engineering and in the workforce more broadly, and is the engineering profession's lead on diversity and inclusion, and would be happy to contribute more information in these areas.

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

32. As with other types of technical advancement, in each sector there are leading organisations, organisations that are fast-followers, those that only change when forced to, and those that refuse to change and die. At this stage of the development of AI, it is mainly pioneering organisations that are active and experimenting. This is generating evidence of how best to generate value that the fast-followers will start to exploit, at which point the use of AI will grow further, with corresponding economic benefits.
33. Successful organisations will be those that treat their data as an asset, partner with organisations in other sectors to get access to different data sources, have a positive engagement with their consumers and partners to ensure ethical concerns are addressed and constantly monitor value versus risks. Limited access to skills could reduce the capacity of organisations to perform each of these activities. Data-centric organisations are fairly flat and laterally integrated, so that a cultural shift may be required as an organization moves away from a hierarchical command and control structure.
34. As mentioned earlier, a small number of the wealthiest companies such as Amazon, Apple, Facebook and Google also own the largest amount of data. This situation has the potential to create even greater disparities between individuals, countries and companies without mechanisms to keep them in check.

Public perception

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

35. There are some commentators that present worst-case scenarios that capture the public's imagination and present challenges to the wider uptake of AI. In the past, misinformation about genetically modified seeds impacted agriculture and lessons should be learned from this. A concerted

public awareness campaign on both the benefits of AI and actions being taken to mitigate the downsides is needed, emphasising how AI in partnership with humans can be more productive.

36. There is a general lack of understanding of the different types of algorithms used in artificial intelligence and the way that they are used. The opportunities and risks associated with the use of algorithms in decision-making depend on the type of algorithm; and understanding of the context in which an algorithm functions will be essential for public acceptance and trust. Similarly, whether an AI system acts as a primary decision maker, or as an important aid and support to a human decision maker, could influence the public's understanding of and engagement with AI.
37. Any discussion with the public will need to focus on specific applications or problems that AI can solve, such as in healthcare or transport, rather than the technology in an abstract sense. It is good that the community is already setting up forums to discuss ethical issues and government and the public should engage with these. In addition, the general public needs a better understanding of concepts such as privacy versus secrecy, how to ensure cybersecurity, plus issues such as the ownership of their data and their rights associated with it.

Industry

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

38. All sectors of the economy stand to benefit from the development and use of artificial intelligence, including advanced manufacturing, built environment, energy, transport, health, aerospace and defence and insurance¹¹³⁵. Many of these applications will be producing large volumes of data generated by the Internet of Things and other novel sources such as social media and crowdsourced data. Many functions that are common across multiple sectors will be impacted by AI. This will include HR, energy efficiency, logistics, business planning and customer support.
39. New business models are emerging across a number of sectors, including built environment, transport, defence and aerospace, where data underpins a service around a product or asset. For example, for smart infrastructure, pervasive monitoring and sensing strategies will generate data that enable the use of preventative maintenance strategies so that maintenance interventions are carried out when needed, rather than after a set number of hours of operation. Reliability will also be improved, as

¹¹³⁵ Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, <http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity>

weaknesses can be detected prior to failure occurring. This could help underpin improvements in infrastructure productivity by contributing to the delivery of new infrastructure, as well as maintaining and operating existing infrastructure at highly resilient levels¹¹³⁶.

40. The transport sector would benefit with the adoption of 'mobility as a service', aimed at providing consumers with relevant choices in their transport solutions. Services would be provided by a multitude of transport operators, co-ordinated by one customer interface organisation that would match individual mobility needs with available transport options. An appropriate legal and regulatory environment is needed to enable this system to work, in which, the UK could take a lead if it so wishes. Machine learning techniques would be essential to the effective and efficient application of 'mobility as a service'.
41. In advanced manufacturing, AI used alongside other technologies such as big data, robotics and the Internet of Things could result in higher performance and more flexible manufacturing systems¹¹³⁷. In the energy sector, such techniques could underpin greater interoperability and flexibility in an energy system that is focused on delivering services to end-users. Defence applications are also moving extremely fast, leading to challenging ethical and governance questions. Further examples are discussed in more detail in a joint report produced by the Academy and the IET in 2015¹¹³⁸.
42. AI technologies will also be used in autonomous systems such as autonomous vehicles, as well as those used in manufacturing, drones, maritime and space systems, and in assistive robots¹¹³⁹. Both physical and non-physical applications of AI will increasingly be employed in collaboration with humans, where AI technology will act in an assistive rather than executive mode. This will necessitate robust human-centred design. A particular system design issue is how best to give the operator the right information to exercise appropriate control. In complex systems the operator may need to be highly trained to deal with decisions handed over by the AI to the human.

How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be

¹¹³⁶ Royal Academy of Engineering response to the National Infrastructure Commission (15 March 2017), *National Infrastructure Commission Technology Study - call for evidence*

¹¹³⁷ Industrial Digitalisation Review - Interim Report (July 2017), Version 3.0

http://industrialdigitalisation.org.uk/wp-content/uploads/2017/07/Interim_Report_Final3_1.pdf

¹¹³⁸ Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, <http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity>

¹¹³⁹ Royal Academy of Engineering (2015), *Innovation in autonomous systems*, <http://www.raeng.org.uk/publications/reports/innovation-in-autonomous-systems>

addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

43. New types of companies have emerged, particularly in the US, whose business models are based on the aggregation of data and provision of cloud services, such as Amazon, Google (now part of Alphabet), Facebook, Microsoft and Apple. These companies hold huge quantities of data and are in a position to compete with traditional engineering sectors for a share of the market in such areas as autonomous cars and smart cities¹¹⁴⁰, as well as with other sectors such as supermarkets, finance and insurance¹¹⁴¹.
44. The major platform vendors may monopolise data, but there are counter examples: for example, Uber has managed to collect the traffic and map data it needs to offer its services and has the potential to rival Google in the autonomous vehicle space.
45. The large data companies are themselves developing artificial intelligence capabilities in-house, in some situations by acquiring AI SME's. Google's acquisition of Deepmind is one such example. The critical mass of expertise in such an organisation is huge, and Deepmind is considerably larger than many university departments¹¹⁴². Large data companies are also competing with SMEs and other types of tech and engineering firms for graduates skilled in data science, AI, robotics and other related areas, and are able to pay large salaries¹¹⁴³. The UK needs to develop a good defence against such competition, although the creation of the Alan Turing Institute has helped counter this, as has the increase in data science courses provided by UK universities¹¹⁴⁴.
46. A better form of data sharing is needed. Data is central to much of the power of AI and the large technology companies have significant

¹¹⁴⁰ Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, <http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity>

¹¹⁴¹ World Economic Forum (August 2017), *Big Tech, Not Fintech, Causing Greatest Disruption to Banking and Insurance* <https://www.weforum.org/press/2017/08/big-tech-not-fintech-causing-greatest-disruption-to-banking-and-insurance>

¹¹⁴² In March 2016, Deepmind employed approximately 140 researchers: <http://www.techworld.com/personal-tech/google-deepmind-what-is-it-how-it-works-should-you-be-scared-3615354/>

¹¹⁴³ A TechNation survey stated that for tech firms in the data management and analytics area, 'barriers to accessing analytical talent are currently preventing companies from reaching their full potential'. TechCity (2016), *TechNation: transforming UK industries*. http://www.techcityuk.com/wp-content/uploads/2016/02/Tech-Nation-2016_FINAL-ONLINE-1.pdf?utm_content=buffer2e58f&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

¹¹⁴⁴ RAEng and IET (2015), *Connecting data: driving productivity and innovation*, <http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity>

advantages based on the data they hold. Finding a way to share this with others would help to level the playing field. Care must be taken, however, to preserve privacy and to comply with the General Data Protection Regulation (GDPR).

47. Trust relies on ensuring that individual, corporate and broader social benefits from data are balanced between stakeholders. There is some evidence that the public are willing to share personal data with companies to get a better service¹¹⁴⁵, but in many instances asymmetries still exist between organisations and consumers so that the organisation has a much better idea of how it can benefit from data than the consumer. There are a number of projects developing platforms^{1146,1147,1148} that allow individuals to control data securely, make it available as they see fit with safeguards and benefit directly from their personal data, responding to the need to rebalance control of data and its benefits. If data is thought of as the 'new oil', the readiness of people to give up personal data without consideration is unwise and a level playing field will not be created without addressing this.

Ethics

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

48. Individuals each have a different definition of what is ethical. Their views are formed from family, culture, education, experience, friends and colleagues. Digital technology in the past has been developed to optimise and automate standard behaviours. The increase in processing power enables these behaviours to become more customised to the situation or personal taste. Thus the variety in an individual's definition of ethical behaviour needs to be taken into account by the system designer to ensure the base premises are reasonable and there is enough flexibility to protect and respect an individual's ethical perspective¹¹⁴⁹.
49. The thinking on ethics in relation to autonomous systems is becoming increasingly well developed. For example, the IEEE Global Initiative for

¹¹⁴⁵ A recent study of travellers' attitudes to intelligent mobility by the Transport Systems Catapult found that 57% of respondents would not mind sharing their personal data in order to get a better service.

¹¹⁴⁶ The Hub-of-All-Things, <http://hubofallthings.com/>.

¹¹⁴⁷ Databox project is developing a privacy-aware data analytics platform to collate, curate, and mediate access to personal data www.databoxproject.uk/

¹¹⁴⁸ Digital Prosumer is developing a platform that allows individuals to take control of the monetisation of their personal data www.digitalprosumer.co.uk

¹¹⁴⁹ Ethics for Big Data and Analytics, <http://www.ibmbigdatahub.com/whitepaper/ethics-big-data-and-analytics>

Ethical Considerations in Artificial Intelligence and Autonomous Systems brings together multiple voices in these communities to identify and find consensus on timely issues. It has recently published a draft document¹¹⁵⁰ on ethical concerns for AI and autonomous systems. Academy Fellows have provided evidence to this initiative and sit on an advisory group. In addition, the British Standards Institute has published a guide to the ethical design and application of robots and robotic systems¹¹⁵¹.

In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

50. Ensuring transparency of algorithmic decision-making is a challenge, particularly for machine learning and self-adaptive systems. Even with techniques to ensure some human supervision, it is already the case that AI is achieving good results, but without humans being able to comprehend how. Issues of governance and accountability will need to be considered in the design and development of these systems. There are many human influences in algorithmic decision-making, including setting criteria choices and optimisation functions that will need to be understood and documented. Software engineering of algorithms will need to introduce mechanisms for logging and providing feedback to allow for greater accountability.
51. The quality of the output from an algorithm depends on the availability, quality and appropriateness of the data that is fed in, as well as the fitness for purpose or 'correctness' of the algorithm itself. Transparency about the data on which the algorithmic decisions are being made is critical to ensure accountability. Good quality metadata is vital for understanding the provenance, quality and timeliness of data.
52. Government, businesses and public bodies will need to consider their use of algorithms in decision-making, consulting widely, and ensuring that mechanisms are in place to detect and address any mistakes, biases or unintended consequences of decisions made. However, the Academy recognises that there are significant implications for government, businesses and public bodies in requiring increased transparency. The regulatory context, commercial constraints, cultural attitudes and the need to protect personal data affect the willingness and ability of organisations to share information to achieve transparency. One example where transparency is required is the key

¹¹⁵⁰ IEEE (December 2016), Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (AI/AS), http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf

¹¹⁵¹ British Standards Institute (April 2016), BS 8611:2016 *Robots and robotic devices - guide to the ethical design and application of robots and robotic systems*.

science journals that require the availability of both data and algorithms before publication.

53. It will be important for data protection safeguards to be built into software and services from the earliest stages of development. In particular, the requirements of the General Data Protection Regulations will have substantial impact, as noted above. There will be requirements for systems with properties that can be checked by regulators or the public without compromising data protection. Mechanisms could include the disclosure of certain key pieces of information, including aggregate results and benchmarks, when communicating algorithmic performance to the public. Further research into effective mechanisms and strong leadership is required to address the evolving intellectual property and legal constraints.
54. In applications such as autonomous vehicles, there are technical challenges around creating fully transparent systems. While it may be possible to understand the internal workings of individual software modules and to specify how these should perform, it may not be possible to map how the performance of individual modules impacts on the overall performance of the system. However, it is the performance of the autonomous vehicle at system level that is the central concern.

The role of the Government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

55. Government should play a role in providing support for research and development, including testbeds and demonstrators as an 'intelligent client' in public-sector procurement and in developing the necessary skills. The Academy welcomes the creation of the Industrial Strategy Challenge Fund that will help support the development of technologies such as artificial intelligence.
56. How AI is used by government, business and public bodies will ultimately determine the level of regulation required for applications of this technology. Future regulations will need to be flexible enough to accommodate different requirements, data types and possible new uses of algorithms in the future, yet ensure protection and consistency in approaches.
57. While the regulatory landscape is developing, government should lead by example by applying standards to its own use of AI, to ensure accountability and help build public trust in use of algorithms. It will be important to consider the protection of personal data, auditability, and liability for harm caused by the use of algorithms.

58. It is important to ensure that regulatory guidance and criteria are developed with sufficient expert input. The Academy stands ready to advise government on regulatory issues, as appropriate. The Academy will be producing a 'challenge paper' on the regulation of autonomous systems. This study aims to support UK industry through better understanding of the key challenges facing UK-based SMEs working on the development and deployment of autonomous systems. It will identify the issues that will drive future regulations for autonomous systems, to inform industrial strategy and the Robotics and Autonomous Systems Sector Deal (the RAS Sector Deal is being led by Fellows on the working group).
59. While the extent of the future use of algorithms in decision making will differ by sector, the Academy believes that an underlying risk is the assumption that algorithms are near-perfect, or will replace humans entirely in all decision making processes. While this might be the case in some sectors, there is a risk that new applications of AI are not being introduced properly or are introduced at the behest of people who do not fully understand how AI works, its limitations, or its potential impact on society. The Academy believes that it is important for government to have an authoritative voice on these matters.
60. The Academy advises careful monitoring by government, alongside businesses and public bodies, where the use of algorithms has a greater scope to introduce or amplify biases or discrimination. This has been noted as a particular concern in financial, recruitment, legal, criminal and education sectors where algorithms may focus on specific metrics, such as age, gender or ethnicity. While this is a significant concern, it also creates the opportunity to remove existing biases by designing systems that are independent of these variables. The issue of biases is emerging as being particularly problematic, although the issue is not with the algorithms *per se*, but rather the nature and labelling of the training dataset.
61. Government needs to recognise that big data is a common good and the fuel of the AI transformation. By ensuring good structures for citizens to control their own data, while making it available for AI applications such as large-scale health applications, government will not only accelerate positive results from AI, but also encourage positive citizenship.
62. Government should also focus on big data sharing for public funded digital projects and for licensed and regulated entities, and to promote data sharing growth as a facilitator for productivity, innovation and investment.

Learning from others

What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

63. The potential to create a 'data-driven' economy is affected by free flows of data across international boundaries, as well as between organisations. The European Commission has identified sharing of data in commercial contexts as a key concern¹¹⁵², and in its working paper discusses the emerging issues around the free flow of data and improved sharing of commercial data, including machine-generated data which are either non-personal in nature or personal data that have been anonymised.
64. A useful benchmark could be provided by the Finnish government's decision to introduce legislation to incentivise the use of 'mobility as a service' in the use of data, machine learning, and the frameworks to share data across traditional 'siloes sectors'.
65. It will also be important to monitor industry consortia such as Open AI and The Partnership for AI¹¹⁵³.
66. The European Parliament is developing its thinking on the legal framework for robotics. A draft report¹¹⁵⁴, authored by a Luxembourg MEP, outlines rules to govern how robots interact with humans and was approved by the European Parliament Committee on Legal Affairs in January 2017. This follows a recent EU project, RoboLaw¹¹⁵⁵, which investigated the ethical and legal principles raised by robotic applications and provided European and national regulators with guidelines to deal with them.

6 September 2017

¹¹⁵² European Commission (January 2017), Commission Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final, <https://ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy>

¹¹⁵³ <https://www.partnershiponai.org>

¹¹⁵⁴ European Parliament Committee on Legal Affairs (May 2016), *Draft report with recommendations to the Commission on Civil Law Rules on Robotics* (2015/2013(INL))

¹¹⁵⁵ EU project RoboLaw (September 2014), *Regulating emerging robotic technologies in Europe: robotics facing law and ethics*, http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf

The Royal College of Radiologists – Written evidence (AIC0146)

Introduction

1. The Royal College of Radiologists (RCR) works with our 10,000 members to improve the standards of practice in the specialties of clinical radiology and clinical oncology.

2. The Royal College of Radiologists welcomes the opportunity to give evidence to this inquiry as clinical radiology and clinical oncology are at the forefront of the use of Artificial Intelligence (AI) in a healthcare setting. For the purposes of this response the RCR is using the definition of AI as the branch of computer science dealing with the simulation of intelligent behaviour in computers. This response focuses on the short and medium term effects of AI.

Summary

- The members and fellows of the RCR are at the forefront of the development of AI in healthcare. We believe that AI may be part of the solution to the current workforce crises in both clinical radiology and clinical oncology but it cannot replace human clinicians.
- The development of AI in healthcare will require a huge amount of patient data. The NHS has these data available but it is crucial all data used are fully anonymised at source (before leaving hospital firewalls), and that patients understand how and why their data are used. The NHS must also get the best deal for any data used by private companies.
- The government has a crucial role to play in the development of AI in healthcare. This includes ensuring there is investment in infrastructure and workforce. Developing ways of sharing the patient data needed to develop AI and ensuring there is a regulatory system that protects both patients and doctors.
- The RCR suggests that AI companies are urged to concentrate upon developing machine learning that can recognise normal plain X-rays of every part of the body with over 99.5% accuracy (e.g. the wrist, the spine, the pelvis etc.). If these could be safely and reliably reported by AI it would be very helpful in assisting with the reporting of normal plain X-rays. This may free up radiologists to report all the abnormal imaging studies and to perform hands-on interventional work and may alleviate some of the workforce pressures on radiology departments. Clinical judgement will still be essential for assessing individual cases. For example a patient who has a negative X-ray may still have significant problem that requires a radiologist to assess what further investigation is needed.
- For AI companies to develop software to recognise normal X-rays from all parts of the body, they need to train their machine learning using huge

amounts of normal plain X-ray data. These data already exist on the PACS (picture archive and communication system) archives in every NHS hospital in the UK, but to safeguard patient privacy they need to be released in a completely anonymised form to a central, independently regulated national database. The RCR is developing a specification stipulating how these normal X-ray data can be anonymised at source before passing through the firewalls of hospitals. Funding is required for the RCR (as an independent expert professional body) to trial and test a prototype of this process, so that a very large national database of normal X-ray studies can safely and quickly be built up for the AI companies to use, for a fee.

- To ensure only quality plain X-ray data (true normals) are contributed to this national database archive, NHS radiologists would need to check, by double reporting, that all X-ray studies submitted are normal. Funding would need to be provided for this.

Evidence

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

3. AI is already in use throughout radiology and oncology and both specialties are likely to be heavily affected by AI over the next 20 years. The RCR does not believe that AI will be able to replace human clinical radiologists or clinical oncologists as AI will not have explanatory power, so cannot investigate cause. AI will also not replace a doctor's judgement, creativity or empathy.

4. The NHS handles over 40 million imaging investigations per year¹¹⁵⁶ and AI has the capability of developing sophisticated scheduling algorithms that dynamically match the needs to the patients, and to the capacity of the resources. All modern radiological imaging is already acquired, viewed, reported and archived electronically, and all modern radiotherapy (excepting skin cancer) is planned on electronic images, using complex suites of computer programmes. For those reasons, AI will make stronger inroads in these areas of healthcare than in others.

¹¹⁵⁶ NHS England, Diagnostic Imaging Dataset Annual Statistical Release 2014/15
<https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2014/11/Annual-Statistical-Release-2014-15-DID-PDF-1.1MB.pdf>

5. For interventional radiology, a subspeciality of radiology, AI may assist with deciding the pathway for each patient (treat, not treat or formal review) for procedures like stroke thrombectomy and angioplasties.

6. Clinical oncology was one of the first specialities routinely to computerise large elements of its work, initially for radiotherapy planning. Modern radiotherapy planning requires the routine use of complex software, often running on dedicated hardware, and is used to produce optimal radiotherapy plans.

7. AI will also have huge potential for breast imaging. There are approximately two million women screened per annum in UK¹¹⁵⁷ and images are read at a rate of 55 per hour. Replacing one of two breast screen readers with AI (all screening mammograms are double read by a radiologist, advanced practitioner or breast physician) would free up a significant amount of time for a service under severe stress due to staffing shortages. Many consultants are now due to retire at the same time, 30 years after the breast screening programme was set up.¹¹⁵⁸

8. As AI develops it is likely that more complex clinical scenarios will be able to be analysed and the interplay between many comorbidities will be understood. However when benefits of therapy are minimal (as in palliative treatment towards the end of life) patients may prefer to explore these value options with a clinician. Key to decision making are the needs and wishes of patients and their families, which may not always align with the outcome of an algorithm.

9. Technical limitations will be one of the biggest factors which may hinder development including computer power and the availability of data. There needs to be investment in computers with parallel processing capability and NHS IT will need to be upgraded to cope with increasing demands. Another limitation will be the need for huge amounts of anonymised patient data for training AI algorithms; we have considered this further in question four. Another limitation will be the availability of anonymised patient data for machine learning. The RCR is working on a specification outlining how plain X-rays can be anonymised before being used by AI companies and as an independent body we could oversee the development of a national database.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

¹¹⁵⁷ Health and Social Care Information Centre, Breast Screening Programme, England Statistics for 2014-15 <http://content.digital.nhs.uk/catalogue/PUB20018>

¹¹⁵⁸ Royal College of Radiologists, The breast imaging and diagnostic workforce in the United Kingdom

https://www.rcr.ac.uk/system/files/publication/field_publication_files/bfcr162_bsbr_survey.pdf

10. There has been some unwarranted hype over AI in the past but there is now a conjunction of software, hardware and data that make further development of AI possible. However, there remains a risk of over-hyping AI in particular areas. In particular, the adoption of AI into clinical practice faces several barriers. Widespread adoption or spread must be done in a measured way that ensures that there is sufficient research, rigorous clinical testing, clinical training and regulation. AI companies should be encouraged to focus on developing software that can recognise normal plain X-rays rather than anything more complex as that is where the most value can be added,

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

11. The public can be prepared for the use of AI in healthcare by fully addressing issues around consent and ownership of data, and by transparent scientific demonstration of the accuracy and reliability of AI in domains where it is to be used. AI systems work best with the very large amount of data available at NHS level rather than at Trust level. Access to these data requires the general public to understand what their data are being used for and have confidence that the development of AI and use of their data are well regulated. Public confidence is likely to depend upon ensuring that the uses of data clearly benefit other patients, and that data are not released to commercial parties without clear safeguards.

12. As AI begins to play a role in clinical radiology and clinical oncology it will be crucial that information is made available about the role that AI has in patient care, and crucially what role AI cannot play, and so will continue to be done by humans. There is a risk that current rhetoric around AI may lead people to believe erroneously that machines can entirely replace humans in this area.

13. The general public can also be prepared for widespread use of AI by the development of education to support the future workforce. In particular higher education needs to ensure that students have the skills to utilise AI techniques. A major challenge at present in the real world applications of AI, such as in healthcare, is bridging the understanding gap between domain experts (clinicians) and AI experts.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

14. All levels of society that use the NHS stand to benefit from AI, given that imaging is indispensable for the whole of modern medicine and surgery, including

cancer pathways and trauma. There is potential for large swathes of routine work to be done by AI which may be disruptive to staff. Within imaging, the workforce of highly skilled and flexible workers will be advantaged as AI has the potential, by “weeding out” and accurately reporting all normal plain X-rays, to free up capacity to allow radiologists and oncologists to concentrate on the more complex studies and work. This approach across imaging will enable radiologists to make use of their high level skills for those patients who need them.

Public perception

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

15. It is essential that the public understand how AI can benefit them, particularly in the context of the delivery of healthcare. The adoption and implementation of AI may improve outcomes, and free up time so that doctors can spend more time with patients and the more complex aspects of their jobs. If the results of certain investigative tests can reliably and accurately be provided by AI, this would mean many patients would wait less time for their test results. The use of AI in such circumstances may make up for workforce vacancies, and can be used day and night 365 days of the year.

16. However to develop AI in clinical radiology and clinical oncology vast amounts of patient data are needed to train the AI software, which means the public must be understood that their data will be safely and thoroughly anonymised, and know how their data will be used. Whilst lots of the early AI work is done in academic institutions, ultimately academic partnerships with industry are required for commercialisation. Patients and the public may well be wary of their data being used for commercial benefit and at the moment this is a grey area.

17. The government should protect intellectual property generated from patient data with careful academic/industry contracts which reward the NHS for its contribution. This could include shares in the profits from commercial products which can be fed back into the NHS/academia, or negotiating free installation and access to the AI products for NHS patients and staff, as well as payment for access to these anonymised data. The concerns about sharing patient data could be reduced by normalising data sharing nationally in the UK e.g. all imaging studies to be used for AI research and training to be rigorously anonymised at source (within hospital firewalls), and then transferred into a national repository, overseen by an independent body such as the RCR. Ideally there would be the implementation of an opt-out system for patients, even when data have been rigorously anonymised. Most importantly there must be transparent systems in place to reassure the public.

Industry

6 What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

18. Medical imaging is perfectly placed to benefit from advances in AI. It is one of the few areas that has large data pools of high quality curated data (in the form of imaging studies and their reports) thus overcoming one of the main hurdles in AI development. The government needs to invest now in medical imaging research if the UK is to play a significant role in AI in the future. The AI technology industries need for data, especially labelled data, puts the NHS in a unique position.

19. The NHS should be mindful that the data AI models rely upon are key. Therefore, negotiations between NHS organisations and AI developers should be carefully monitored to avoid handing over data for little, or no, reward. At present, the tendency is to release the data to developers. However, this approach is not without risk so there should be the development of centralised AI resources available within the NHS. Such an NHS AI Institute could be relatively small, and yet provide shared resources and expertise across many domains. Patients may feel more comfortable with their data being used within the NHS rather than outside and have the process of anonymising data overseen by the appropriate body, in the case of medical imaging this should be the Royal College of Radiologists

20. The risk with AI is that a small number of industry leaders, once ahead of their competitors, will develop an unassailable advantage, due to the nature of iterative reinforcement learning. For example, an imaging product becomes useful and is adopted. New images are passed through the product as it is being used and as a result the product improves as it 'sees' more examples. Given the global nature of the AI industry, the UK must invest now in order to maintain a foothold in this industry in the future. There is potential for the NHS to achieve substantial strides in medical AI given that it is the world's largest single-payer healthcare provider, and continues to enjoy public support.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

21. Any transfer of patient data between the NHS and private companies must get value for money for the NHS. The NHS has invested significant resource in developing banks of data that are now being used by technology companies. It is vital that this effort is recognised and appropriately rewarded by private companies who will profit from it. Given the complexity of the area, and the potential for network effects, there should be some consideration to the idea that

an NHS AI Institute should supervise such deals. The risk is that otherwise commercial entities may be able to cherry-pick deals with individual NHS Trusts.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

22. There are ethical issues around consent for data to be used by AI and also consent of patients for AI to be used in their treatment. We have explored some of these in issues in questions three and five above. If the data are thoroughly anonymised before passing through hospital firewalls, this should provide reassurance to patients

23. Private companies using NHS data must ensure that they abide by the same ethical standards as the NHS for studies. This includes ensuring applications to ethics committees for use of data are completed. Security of patient data is also important; one way of overcoming this could be for data to be held by the NHS in a centralised resource. However, there are wider ethical issues around trust and liability in AI-enabled healthcare. There remain open questions about when failure to use an AI system would be negligent, or how errors in a combined human/ AI system would be addressed.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

24. AI has the capacity to process information to degrees of complexity that far exceeds the comprehension of the human mind. Much work is being done to decipher 'black box' algorithms as this enables improvements to be made, and this will shed some light on the process but will not be comprehensive. Indeed test-validation of complex algorithms may be unfeasible in practice. More generally, however, there is a gap between the validation of algorithms and the validation of their implementation clinically. Although there are guidelines on software engineering (e.g. as used in the Space Shuttle), these are rarely followed in current AI software. This should be considered in regulation.

The role of the Government

10 What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Data and infrastructure

25. For AI to be a success there needs to be upfront investment in IT systems. Currently effective IT is not available and this will slow the implementation of AI. Such investment will mean improved decision making which will have a cost saving.

26. NHS data should be viewed as a key strategic resource, modern AI is extremely data dependent, and a licensing model that allows collaboration with commercial entities while ensuring that benefits are returned to the NHS, is vital. The nature of AI-based systems, and their dependence on large amounts of data, creates the possibility of creating a 'winner-takes all' situation where once a system has established a lead, it becomes impossible for competitors to overcome its advantages. The organisation of such systems also has the potential to reinforce economic inequality, with large gains flowing to relatively small groups of people/companies. It seems unlikely that using large amounts of NHS patient data to generate profits for small numbers of companies would be acceptable to the public.

27. There should be consideration of developing some centralised, sharable resource of AI computing, especially for building models over large NHS datasets. Such infrastructure could be virtual but it would provide a central point of contact and expertise to provide overlap between the clinical and AI worlds, with appropriate safeguards. The RCR can assist in ensuring that data used is fully anonymised.

Regulation and governance

28. The governance of the use of AI in healthcare is unclear, and problematic. At present, clinicians are held responsible for deciding to use, and interpret the results of most AI tools. Such tools include auto-contouring software in radiotherapy planning or decision support software for breast cancer chemotherapy.

29. However, as AI becomes more autonomous, then governance becomes more difficult. It is not clear at what point failure to use an AI system would become negligent. Linked to this is the relatively poor development and documentation process for much software development. Although there are methodologies for writing safety-critical software, these are currently rarely used in AI systems. In addition, the tendency for software to be provided as a service, rather than as a downloadable product, makes it easy for the provider incrementally to upgrade the product. While this has many advantages, it may mean that the AI service is not stable over time. Auditing the results of such a service is therefore challenging.

30. Without a clear governance framework for managing the risk associated with using AI tools, their introduction is likely to be slow, and haphazard. This framework would need to include multiple elements, including the data used to build such tools, software engineering practices and updates, and a clear process

for understanding and assigning risk. This discussion will need to involve a range of partners, including regulators and indemnity organisations

31. Regulation of AI in healthcare will be crucial: not just the regulation of software itself by Medicines and Healthcare products Regulatory Agency (MHRA) but the impact AI may have on the regulation of healthcare professionals. If a doctor uses a treatment recommended by an algorithm where do the professional responsibilities of the doctor end? A regulatory framework should consider such issues.

Liability

32. Legal liability is often stated as a major societal hurdle to overcome before widespread adoption of AI becomes a reality. If mistakes are made in the course of treatment or diagnoses where AI is used there will be a need for robust legislation to cover liability.

Investment

33. Given the amount of AI talent, both commercially and academically, and the presence of large datasets (especially healthcare-related) there is a potential for the UK to be a world leader in AI. The government needs to ensure it seizes this opportunity by ensuring that there is adequate investment and support for AI. This should include providing funding for independent bodies like the RCR to oversee the development of systems of anonymise patient data.

34. The government needs to clarify immediately the status of foreign workers as the elite in this field are few in number, and the UK needs to ensure that it is a welcoming place to work. The barriers in this sector, real, or perceived, are damaging to the workforce

35. The National Institute for Health Research (NIHR) should consider setting up an NIHR AI Bioresource similar to the approach taken to genomics, with the possibility of it being spun out into a private company equivalent to Genomic England.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

36. The American College of Radiology has established a [Data Science Institute](#) to guide the appropriate development and implementation of AI tools to help radiologists improve medical imaging care.

Dr Nicola Strickland
President

Royal College of Radiologists

6 September 2017

The Royal Society – Written evidence (AIC0168)

6 September 2017

Submission to the House of Lords Select Committee on Artificial Intelligence

Summary

Recent years have seen exciting advances in machine learning, which have raised its capabilities across a suite of applications. Machine learning is a branch of artificial intelligence (AI) that uses algorithms to learn directly from examples, data, and experience. It is the form of AI that many people now interact with on a daily basis, and which poses policy questions over the next ten years.

Increasing data availability has allowed machine learning systems to be trained on a large pool of examples, while increasing computer processing power has supported the analytical capabilities of these systems. Within the field itself there have also been algorithmic advances, which have given machine learning greater power.

As a result of these advances, systems which only a few years ago performed at noticeably below-human levels can now outperform humans at some specific tasks. Many people now interact with systems based on machine learning every day, for example in image recognition systems, such as those used on social media; voice recognition systems, used by virtual personal assistants; and recommender systems, such as those used by online retailers.

In addition to these current applications, the field also holds significant future potential; further applications of machine learning are already in development in a diverse range of fields, including healthcare, education, transport, and more. The social and economic opportunities which follow are significant. There is a vast range of potential benefits from further uptake of machine learning across industry sectors, and the economic effects of this technology could play a central role in helping to address the UK's productivity gap.

While offering potential for new businesses or areas of the UK economy to thrive, the disruptive potential of machine learning brings with it challenges for society, and questions about its social consequences. Some of these challenges relate to the way in which new uses of data reframe traditional concepts of, for example, privacy or consent, while others relate to how people interact with machine learning systems. Careful stewardship will be needed to ensure that the productivity dividend from machine learning benefits all in society, and there is an opportunity now to shape how this technology develops so that its benefits can be shared across society.

The UK can maintain its leading position in developing machine learning and AI:

- By supporting the development of skills at every level, through interventions from schools to universities, and support for further PhD places in this area of research;

- By helping create opportunities to use machine learning through support for business and research, through both Industrial Strategy and UKRI;
- By creating a data environment that supports machine learning, through further action on open data and standards;
- By leading a societal debate about machine learning and AI, and supporting researchers in public engagement on these topics; and
- Through an enabling governance environment that creates a new framework for the governance of data in the 21st century. Beyond this, it is not appropriate to establish an overarching governance framework for machine learning. The issues involved differ greatly across the many sectors in which machine learning is applied (including healthcare, transport, education, scientific research, finance, retail, and public policy), and are thus better handled on a sector-by-sector basis.

In addition to these near-term policy measures, machine learning raises further questions for the future:

- What impact will machine learning have on work, and how should we manage this?
- How can we ensure the field develops in a direction that provides broad social benefits?
- How can we make sure the benefits of machine learning are shared?

The Royal Society would welcome the opportunity to discuss these issues further with the Committee.

1. The pace of technological change

Evolution of machine learning and current capabilities

1.1 The Royal Society is the UK's national academy of science. It is a self-governing Fellowship of many of the world's most distinguished scientists working in academia, charities, industry, and public service. Its fundamental purpose is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity. Within this its strategic priorities include providing scientific advice for policy, and education and public engagement.

1.2 The Royal Society's report *Machine learning: the power and promise of computers that learn by example* sets out the potential of machine learning over the next five to ten years, and the actions necessary to allow society to benefit fully from the development of this technology.¹¹⁵⁹ The findings of this report, and

¹¹⁵⁹ Machine learning is the technology that allows computer systems to learn from examples, data, and experience. If the broad field of artificial intelligence (AI) is the science of making machines smart, then machine learning is a technology that allows computers to perform specific tasks intelligently, by learning from examples. These systems can therefore carry out complex processes by learning from data, rather than following pre-programmed rules.

the subsequent report in collaboration with the British Academy *Data management and use: Governance in the 21st century*, offer insights into the societal and ethical issues raised by AI, public perception, the challenges and opportunities for AI in industry, the challenges of data governance in the 21st century, and the role of the Government in the development and use of AI.

1.3 Recent years have seen significant advances in the capabilities of machine learning. Many people now interact with machine learning-driven systems on a daily basis¹¹⁶⁰. In addition to these current applications, the field also holds significant future potential; further applications of machine learning are already in development in a diverse range of fields, including healthcare, education, transport, and more.

1.4 As a result of these recent advances, machine learning is already able to achieve a higher level of performance than people in some specific areas or tasks. For other tasks, human performance remains much better than that of machine learning systems. For example, recent advances in image recognition have made these systems more accurate than ever before. In one image labelling challenge, the accuracy of machine learning has increased from 72% in 2010, to 96% in 2015, surpassing human accuracy at this task.¹¹⁶¹ However, human-level performance at visual recognition in more general terms remains considerably higher than these systems can achieve.

1.5 In addition to algorithmic advances, which have increased its technical capabilities, the progress made in this field owes much to the increasing availability of data and of computing power:

- If one thinks of machine learning systems as algorithms that learn from examples, there has been an explosion in some areas in the last few years in the sets of available examples on which they can be trained.
- Many advanced machine learning systems require massive computing power in order to support their analytical capabilities, and the processing power of computers has vastly increased in recent decades.¹¹⁶²

1.6 While the significant progress made in recent years has enabled many impressive advances, machine learning remains subject to a number of limitations on its use. For example: some approaches to machine learning rely on access to large amounts of labelled training data; it is difficult to develop

¹¹⁶⁰ For example in image recognition systems, such as those used to tag photos on social media; in voice recognition systems, such as those used by virtual personal assistants; and in recommender systems, such as those used by online retailers

¹¹⁶¹ The Economist. 2016 From not working to neural networking. See <http://www.economist.com/news/specialreport/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not> (accessed 22 March 2017).

¹¹⁶² Moore G. 1965 Cramming more components onto integrated circuits. *Electronics* 38, 114.

systems with contextual understanding of a problem (“common sense”); it is difficult to transfer learning from one problem domain to another; and so on. These limitations are discussed in more detail in the Society’s Machine Learning report. Technical advances may help directly address these limitations.

Areas for action to influence the development of machine learning in the next five to ten years

1.7 To realise the potential of machine learning over the next five to ten years, action is needed in four key areas, outlined below. Further detail about each of these areas is available in the Society’s Machine Learning report.

1.8 Supporting **the development of machine learning requires an amenable data environment, based on:**

- **continued Government open data efforts;**
- **new models of data sharing that respect privacy and enable carefully-managed access to certain datasets, for example from the NHS;**
- **resources for data management within research funding; and**
- **extending the lifecycle of open data through open data standards.**

1.9 **As machine learning systems become ubiquitous, and a more significant part of people’s lives and livelihoods, three skills needs follow.**

- **A basic understanding of the use of data and these systems will become an important tool required by people of all ages and backgrounds. Introducing key concepts in machine learning as well as some of the key social and ethical issues at school can help cultivate these skills.**
- **New mechanisms are needed to create a pool of informed users or practitioners. This requires adjusting university course provision in disciplines such as law, healthcare, and finance. Additionally, a new funded programme of Masters courses may also help to increase the number of informed users of machine learning within specific sectors which could benefit from the technology.**
- **There is already high demand for people with advanced skills, and additional resources to increase this talent pool are critically needed. These resources include increasing provision for training PhD students, and creating mechanisms to recruit and retain outstanding research leaders in machine learning in the academic sector.**

1.10 Businesses of all sizes across sectors need to have access to appropriate support that helps them to understand the value of data and machine learning to their operations. Such support includes:

- **access to talent;**
- **advice via government support mechanisms for business; and**
- **measures to promote machine learning through the industrial strategy.**

1.11 While offering potential for new businesses or areas of the UK economy to thrive, the disruptive potential of machine learning brings with it challenges for society, and questions about its social consequences.

- **While it is not appropriate to set up governance structures for machine learning per se, governance surrounding the use of data requires a new framework to keep pace with the challenges in the 21st century.**
- **Continuous engagement between machine learning researchers and the public will be important as the field develops. This should be complemented by relevant ethics training for machine learning postgraduates.**

1.12 Machine learning is a vibrant field of research, with a range of exciting areas for further development across different methods and applications. There is a collection of specific research questions where progress would directly address potential public concerns around machine learning, or constraints on its wider use, for example in interpretability, fairness, and verification and validation. These challenges are outlined in more detail in the report, and **support for research in these areas is needed to help ensure continued public confidence in the deployment of machine learning systems.**

2 Impact on society

Human capital, and building skills at every level

2.1 To thrive in an environment shaped by widespread use of by machine learning, and in which machine learning is a key tool for daily activities and work, citizens will require data literacy skills, which enable them to use and interact with data, and an understanding of the strengths and weaknesses of technologies such as machine learning. These skills – and the policy measures to support them – are considered in more detail in Chapter 4 of the Society's Machine Learning report.

Opening further debates

2.2 In addition to the near-term policy measures set out in paragraphs 1.8 to 1.15, machine learning raises further questions for the future:

- What impact will machine learning have on work, and how should we manage this?

Common ground on the nature, scope, and scale of the impact of AI on employment is difficult to establish: different AI technologies can be put to use in different ways to automate different tasks in different fields and to different timelines, in addition to creating new types of work or opportunities for human-machine collaboration.

Through the varying estimates of jobs lost or created, tasks automated, or productivity increased, there remains a clear message: machine learning will have a significant impact on the way we work, and its effects will be felt across the economy.

What is less clear, however, is whether the changes that arise as a result of the impact of machine learning will be like-for-like, with new tasks created, or whether certain tasks, roles, or people will be displaced.

There will be an enduring question about how machine learning and AI change the way we all work. The Royal Society will continue to explore these questions.

- How can we ensure the field develops in a direction that provides broad social benefits?

It is highly likely that it is not just machine learning, but machine learning alongside other data-based techniques and advances, such as those in robotics, will be disruptive. In attempting to understand the current landscape and interpret how decisions about machine learning made today might affect its future, taking account of the growing body of insight into how emerging technologies are viewed and used as they move from novel to mainstream may be helpful. The nature, scale, and duration of this disruption will depend on the social, political, ethical, and legal environments in which these technologies evolve.

- How can we make sure the benefits of machine learning are widely shared?

Previous major waves of technological change, including the industrial revolution, the use of electricity, and the development of electronics, have also been characterised by productivity increases. There have been benefits across society through raised living standards and wellbeing, as well as substantial financial benefits to a small subset of individuals or corporations. There have also been changes in the work environment.

In the same way, there will be a 'productivity dividend' generated by machine learning, in parallel with changes to the world of work and other aspects of people's lives. What is not clear is how the productivity dividend will be shared and who the major beneficiaries will be.

At this early stage of the cycle it may be possible for society to shape the way the productivity dividend is shared, for example by engaging industry in discussions about how the demand for new skills can be met (and funded), and other policy responses to ensure there are not groups of people left behind as a result of social changes to which this technology contributes.

Thinking about how the benefits of machine learning can be shared by all is a key challenge for all of society. **Society needs to give urgent consideration to the ways in which the benefits from machine learning can be shared across society.**

Cybersecurity

2.3 As devices, such as phones or household appliances, becoming increasingly 'smart', with the ability to collect and analyse data and communicate with other devices or control units (the so-called 'Internet of Things'), they will be able to respond more intelligently to our needs.

2.4 There are valid worries about the security of these devices, and what might happen if a malicious user (or 'virus' type algorithm) were to gain control of an increasingly large and inter-connected part of our environment.

2.5 The Royal Society's report *Progress and research in cybersecurity* noted how technical and social change required new approaches to cybersecurity, and that progressing these would require substantial research and development. Its recommendations noted the importance of:

- **Trust: Governments must commit to preserving the robustness of encryption, including end-to-end encryption, and promoting its widespread use.** Encryption is a foundational security technology that is needed to build user trust, improve security standards and fully realise the benefits of digital systems.
- **Resilience: Government should commission an independent review of the UK's future cybersecurity needs, focused on the institutional structures needed to support resilient and trustworthy digital systems in the medium and longer term.** A self-improving, resilient digital environment will need to be guided and governed by institutions that are transparent, expert and have a clear and widely-understood remit.
- **Research: A step change in cybersecurity research and practice should be pursued; it will require a new approach to research, focused on identifying ambitious high-level goals and enabling excellent researchers to pursue those ambitions.** This would build on the UK's existing strengths in many aspects of cybersecurity research and ultimately help build a resilient and trusted digital sector based on excellent research and world-class expertise.
- **Translation: The UK should promote a free and unencumbered flow of cybersecurity ideas from research to practical use and support approaches that have public benefits beyond their short term financial return.** The unanticipated nature of future cyber threats means that a diverse set of cybersecurity ideas and approaches will be needed to

build resilience and adaptivity. Many of the most valuable ideas will have broad security benefits for the public, beyond any direct financial returns.

3 Public perception

3.1 From the start of its machine learning project, the Royal Society has been engaging with the public to find out their existing attitudes towards machine learning. A public dialogue exercise on machine learning, carried out during 2016 in conjunction with Ipsos MORI, was a key part of this process.

3.2 Public views on machine learning depended on the context in which the machine learning algorithms were being applied, and differed very substantially across different application areas. Attitudes to the technology of machine learning itself were largely neutral. In evaluating the desirability of machine learning in different applications, participants took a broadly pragmatic approach, assessing the technology on the basis of:

- the perceived intention of those using the technology;
- who the beneficiaries would be;
- how necessary it was to use machine learning, rather than other approaches;
- whether there were activities that felt clearly inappropriate; and
- whether a human is involved in decision-making.

Accuracy and the consequences of errors were also key considerations.

3.3 At its core, this dialogue exercise showed that the public do not have a single view on machine learning. Attitudes towards this technology – whether positive or negative – depend on the circumstances or application in which it is being used. This context is key, as the nature or extent of public concerns, and the perception of potential opportunities, are linked to the application being considered.

3.4 Fundamentally, the issues raised in these public dialogues related less to whether machine learning technology should be implemented, but how best to exploit it for the public good. Such judgements were made more easily in terms of specific applications, than in terms of broad, abstract principles. The public saw most to be gained where machine learning could be used to augment human abilities, or do things humans cannot, for example providing advanced analysis.

3.5 Continued engagement between machine learning researchers and the public is needed: those working in machine learning should be aware of public attitudes to the technology they are advancing, and largescale programmes in this area should include funding for public engagement activities by researchers. Government could further support this through its public engagement framework.

3.6 The Royal Society will be building on this work in coming years by creating spaces for further dialogue and interaction.

4 Industry

Emerging applications of machine learning

4.1 As the field develops further, machine learning shows promise of supporting potentially transformative advances in a range of areas, and the social and economic opportunities which follow are significant. In healthcare, machine learning is creating systems that can help doctors give more accurate or effective diagnoses for certain conditions. In transport, it is supporting the development of autonomous vehicles, and helping to make existing transport networks more efficient. For public services it has the potential to target support more effectively to those in need, or to tailor services to users. And in science, machine learning is helping to make sense of the vast amount of data available to researchers today, offering new insights into biology, physics, the social sciences, and more.

4.2 There is a vast range of potential benefits from further uptake of machine learning across industry sectors, and the economic effects of this technology could play a central role in helping to address the UK's productivity gap. Some of the potential applications across sectors are considered in more detail in Chapter 2 of the Society's Machine Learning report.

Creating value from machine learning

4.3 To meet the demand for machine learning across industry sectors, the UK will need to support an active machine learning sector that capitalises on the UK's strengths in this area, and its relative international competitive advantages.¹¹⁶³

4.4 In recent years, the UK's machine learning community has demonstrated its strength in supporting start-ups. On the one hand, the recent acquisitions of DeepMind, VocalIQ, Swiftkey, and Magic Pony, by Google, Apple, Microsoft, and Twitter respectively, point to the success of UK startups in this sector. On the other, they reinforce the sense that the UK environment and investor expectations encourage the sale of technologies and technology companies before they have reached their full potential. **Strategic consideration should**

¹¹⁶³ The issues introduced below are each considered in more detail in Chapter 4 of the Society's Machine Learning report.

also be given to the right long term approach to maximising value from entrepreneurial activity in this space.

4.5 Both machine learning start-ups and start-up companies in other areas who wish to benefit from machine learning expertise face a number of challenges. For example:

- Perhaps the most critical issue for machine learning start-up companies is human capital and talent. That talent is free to work or migrate anywhere in the world, and extensive global demand considerably exceeds supply. Success of UK start-up companies based around machine learning thus depends on the appeal of the UK, generally, and specifically the ecosystem for start-up companies, to allow them to attract and retain the best and brightest employees.
- Both the development and the training of machine learning algorithms can be extremely computationally intensive. Access to, and funds for, extensive computing can thus provide a competitive advantage.
- While many typical university spin-out companies would be based around a specific technological discovery, this standard model may fit less well for machine learning spin-outs. There may not be any IP per se to be licensed or transferred into a machine learning spinout but rather know-how on the part of the academic founders that is central to the new business.
- One direct way in which governments can potentially help start-up companies, where appropriate and allowable, is through their procurement processes. Government contracts help early-stage companies in several ways: they provide a source of income; they give the company the direct experience of engaging with customers, which provides important feedback for their developing market offering; and they act as external recognition of the company's product.

4.6 An approach to capitalise on the UK's strengths in this area will need to draw from coordinated government action to support machine learning at all levels, including marshalling government procurement practices in a way that treats machine learning as a priority area for investment, supporting UK-based machine learning businesses, and recognising the significance of machine learning in government support for business.

- **Government support for business should be able to provide advice and guidance about how to make best use of data, and organisations such as Growth Hubs or the Knowledge Transfer Network should ensure their business advisers are sufficiently informed about the value of data as business infrastructure to be able to provide guidance for businesses about, for example, the value of machine learning.**
- **The Department for Business, Energy and Industrial Strategy (BEIS) should review support networks for small businesses to ensure they are able to provide advice and guidance about how to make use of machine learning, or to effectively support businesses**

offering machine learning products. This includes public-sector procurement processes, and the effectiveness of support for businesses using machine learning should be considered as part of the Government’s review of the Small Business Research Initiative.

- **Government’s proposal that robotics and AI could be an area for early attention by the Industrial Strategy Challenge Fund is welcome. Machine learning should be considered a key technology in this field, and one which holds significant promise for a range of industry sectors. UK Research and Innovation (UKRI) should ensure machine learning is noted as a key technology in the Robotics and AI Challenge area.**

Maintaining a leading role in academic research and teaching

4.7 There is already high demand for people with advanced skills in machine learning. Specialists in the field are highly sought after in the global market, and can command salaries accordingly. This creates a challenge for academic research in machine learning; there is a growing range of companies which – recognising the value of machine learning to their business – are voracious consumers of talented machine learning researchers, often offering very attractive packages. A continuing strong presence in machine learning within the university sector will be essential to ensure the delivery of training in machine learning as part of undergraduate and postgraduate taught degrees, and also of the next generation of research leaders.

4.8 If the UK is to remain at the forefront of developing this field, then further action is required to help cultivate advanced skills in machine learning, to support both academic and industrial advances. An effective public sector research funding environment should support this, taking a role in driving machine learning research which complements that within industry.

5 Ethics

Machine learning in society

5.1 As it enhances our analytical capabilities, machine learning challenges key governance concepts such as privacy and consent, shines new light on risks such as statistical stereotyping and raises novel issues around interpretability, verification, and robustness. The Society’s Machine Learning report explores the following ethical implications of machine learning and AI in further detail:

- Use of data, privacy, and consent;
- Fairness and statistical stereotyping;
- Interpretability and transparency;
- Responsibility and accountability.

5.2 There are ethical questions around some applications of machine learning, such as whether algorithms need to be interpretable in particular use cases,

when humans should be involved in decision processes, and when algorithms should be held to a higher standard of accuracy or interpretability than human decision-makers. The answers to these questions will vary with the application area. This application-specificity is key when considering machine learning, and was a core message in the Society's public dialogues: some applications may require regulation to ensure public confidence, while others will be non-controversial. Some may be dealt with adequately via existing mechanisms.

5.3 The specific question of when a lack of transparency ("black box") may be acceptable is an example of a question which should be addressed in a context-specific way. Both formal and informal human decision making is itself often far from transparent. While there may be contexts in which it is appropriate to insist on higher standards for algorithmic decision making, this should not be a uniform stipulation. With current technologies, insisting on transparency will often involve a decrease in performance. Our public dialogue process showed that at least in some application areas (including healthcare for example), many people would prefer a more accurate algorithm which was not transparent to one which was transparent but less accurate.

Tensions in data use

5.3 Many of the choices that society will need to make as data-enabled technologies become more widely adopted can be thought of as a series of pervasive tensions, which illustrate the kinds of dilemmas that society will need to navigate.

- Using data relating to individuals and communities to provide more effective public and commercial services, while not limiting the information and choices available.
- Promoting and distributing the benefits of data use fairly across society while ensuring acceptable levels of risk for individuals and communities.
- Promoting and encouraging innovation, while ensuring that it addresses societal needs and reflects public interest.
- Making use of the data gathered through daily interaction to provide more efficient services and security, while respecting the presence of spheres of privacy.
- Providing ways to exercise reasonable control over data relating to individuals while encouraging data sharing for private and public benefit.
- Incentivising innovative uses of data while ensuring that such data can be traded and transferred in mutually beneficial ways.
- Making the most of the ability of algorithms to provide accurate outcomes beyond the human ability while ensuring appropriate levels of interpretability and transparency, and allowing for systems of accountability to be put in place.

- Facilitating debate and engagement while ensuring that such debate is meaningful (reciprocal, has the capacity to shape policy and includes an open and accessible articulation of competing values at stake).

This list will undoubtedly evolve in unpredictable and unanticipated ways. What can be stated with certainty is that the use of data-enabled technologies will continue to give rise to situations where important choices will need to be made. These choices will usually resist simple maximisation or optimisation, though technological developments may change the nature of these tensions in future.

6 The role of the Government

Data management and use

6.1 It is not appropriate to set up governance structures for machine learning per se. While there may be specific questions about the use of machine learning in specific circumstances, these should be handled in a sector-specific way, rather than via an overarching framework for all uses of machine learning; some sectors may have existing regulatory mechanisms that can manage, while in others there may not be these existing systems.

6.2 There are governance issues surrounding the use of data, including those concerning the sources of data, and the purposes for which it is used. For this, a new framework for data governance – one that can keep pace with the challenge of data governance in the 21st century – is necessary to address the novel questions arising in the new digital environment.

6.3 The Royal Society and British Academy's study on data management and use concluded that two types of response are necessary:

- First, **a renewed governance framework needs to ensure trustworthiness and trust in the management and use of data as a whole. This need can be met through a set of high-level principles that would cut across any data governance attempt, helping to ensure confidence in the whole system.** These are not principles to fix definitively in law, but to visibly sit behind all attempts at data governance across sectors, from regulation to voluntary standards.
- Second, **it is necessary to create a body to steward the evolution of the data governance landscape as a whole.** Such a body would not duplicate the efforts of any existing body. Rather, it would seek to ensure that the complete suite of functions essential to governance and to the application of the high-level governance principles is being carried out across the diverse set of public and private data governance actors. These functions would include activities to anticipate future challenges and to

make connections between areas of data governance. Because many types of data management – or technologies making use of data – have significant or contested social values embedded within them, such a body would need strong capacities for public engagement, deliberation and debate.

Maintaining the UK's leading role in developing machine learning

6.4 The UK has a strong history of leadership in machine learning. From early thinkers in the field, through to recent commercial successes, the UK has supported excellence in research, which has contributed to the recent advances in machine learning that promise such potential. These strengths in research and development mean that the UK is well placed to take a leading role in the future development of machine learning. Ensuring the best possible environment for the safe and rapid deployment of machine learning will be essential for enhancing the UK's economic growth, wellbeing, and security, and for unlocking the value of 'big data'. **Action by Government in key areas – shaping the data landscape, building skills, supporting business, and advancing research – can help create an enabling environment that allows the UK to continue to play a leading role in this field.**

7 Learning from others

7.1 Science is a global endeavour. A major reason for the success of UK science and technology is that it has been open and welcoming to the best talent from around the world. Today, 30% of our academic research staff are from abroad and a third of UK start-ups were founded by non-UK nationals. The UK is second only to the US as a destination for global talent. Their presence ensures that we remain first-rate, and importantly, produces a first-rate environment for training home-grown talent.¹¹⁶⁴

7.2 Decisions over how best to conduct research and safely exploit new applications are shared by scientists and governments across the world. Where these may have global impacts, there is value in developing a consistent approach. Due to its world-class research base, the UK's researchers and institutions are well-placed to inform international policy that governs research.¹¹⁶⁵

¹¹⁶⁴ Royal Society (2016) President's address <https://royalsociety.org/news/2016/11/president-anniversary-address/>

¹¹⁶⁵ Royal Society (2016) *UK research and the European Union: The Role of EU regulation and policy on UK research* <https://royalsociety.org/~media/policy/projects/eu-uk-funding/phase-3/EU-regulation-and-policy-in-governing-UK-research.pdf>

7.3 International debates on the development of AI are taking place across the world,¹¹⁶⁶ and The Royal Society is shaping these debates. For example:

- In January 2017 the Society and the US National Academy of Sciences brought together leading figures in AI and machine learning to discuss the development of these technologies, and their associated ethical and societal questions.¹¹⁶⁷
- In October 2017, The Society will be speaking about machine learning at the STS Forum.

6 September 2017

¹¹⁶⁶ See, for example, reports by the Obama administration's White House, the US National Academy of Sciences, EU Parliament, and French Parliament.

¹¹⁶⁷ Material from this meeting can be viewed online at:

<http://www.nasonline.org/programs/sackler-forum/frontiers-machine-learning.html>

The Royal Statistical Society – Written evidence (AIC0218)

Written evidence from the Royal Statistical Society to House of Lords Select Committee inquiry on the implications of Artificial Intelligence

The Royal Statistical Society (RSS) is a learned society and professional body for statisticians and data analysts, and a charity which promotes statistics for the public good. We have around 8000 members in the UK and around the world, and our key strategic goals support the use of statistics and data in the public interest, education for statistical literacy, strengthening the discipline of statistics, and development of the skills of statistical professions. The RSS has a Data Science Section, and hosts a network focused on machine learning.

Summary

There is huge potential for society to benefit from the application and development of artificial intelligence (AI), data science and statistics. There is clearly scope for further and future growth in AI, driven by new technologies and applications, and by exponential growth in the volume and variety of digital data. The government's priorities in its industrial strategy green paper suggested to us some important avenues for the future development of AI, and we are pleased to develop the following recommendations with regard to your independent Inquiry.

- To increase the opportunities presented by AI, we need to strengthen recruitment into AI and related fields. There is a need to strengthen the UK's skills base for this, particularly our nation's quantitative skills. We support the recommendation by Professor Sir Adrian Smith of a study of the long-term implications of the rise of data science for education and skills, allied with support for greater participation in mathematical education.
- While increasing national investment in science and research, we need to shore up research capabilities for AI and related disciplines. We look to the new UKRI to prioritise capabilities in statistics, data science and AI, as well as fundamental mathematical research, and to break down silos between research councils for multidisciplinary work in these areas. The Alan Turing Institute's activities could also expand to support a much wider network of research and teaching in support of data science across the UK.
- Data infrastructure should be a priority as it is a basis for the data economy. Arrangements for data access, local data and open data can be strengthened, and data protection regulation – including the adoption of the EU's General Data Protection Regulation and our subsequent approach within domestic legislation – will be crucial to AI and related fields.

With the growth of digital data there is also, however, growing scope for failure in how such data are accessed and used. To help establish a trusted basis for action across all of the above activities, we recommend:

- Stronger deliberation on questions relating to data ethics. We are working with the Nuffield Foundation and others to create a new Convention on Data Ethics which will help to explore the new ethical challenges posed by AI, machine learning and data science.
- Support for public engagement with regard to data science and artificial intelligence. Without effective public deliberation, conclusions cannot readily be drawn on public views, particularly about the uses of personal data and the desired benefits of such uses.
- Development of professional standards for data science, and application of ethical principles. Professional bodies should also take a lead on developing standards, and the RSS's Data Science Section is willing to play its part in this.

Evidence in full

1. Public perception

1.1. Data and algorithms are fundamental to our economy and to people's day to day lives. For example, global consultancy firm McKinsey estimate that \$2.8 trillion was contributed to global GDP from data flows in 2014 (compared to \$2.7 trillion from flows of goods).¹¹⁶⁸ The volume and variety of data that could be made available for analysis has exponentially grown, and developments in statistics, data science and artificial intelligence will be essential to make good uses of these new sources of data.

1.2. In an age when data-driven technologies and industries are growing at a furious rate, one key limit placed on public understanding and participation in AI will be access to the appropriate level of education, skills and experience. In their reporting on the global economy, McKinsey have recognised that very few of the countries participating in global data flows have adequately supported their workers and communities to participate as the economy changes, and they strongly recommend developing clearer paths to new roles.¹¹⁶⁹ For the UK to be a global leader in AI and machine learning, we need a stronger quantitative skills base. The RSS supports, in our *Data Manifesto*, the strengthening of education and training pathways, to ensure that preparation for statistical and data literacy is widened in school, and continues in colleges and universities and into the world of work.

¹¹⁶⁸ Box 3. Valuing cross-border data flows' in Manyika, J. Lund, S. Bughin, J. Woetzel, J. Stamenov, K. Dhingra, D. (2016) *Digital globalization: The new era of data flows* [PDF], McKinsey Global Institute.
<http://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Digital%20globalization%20The%20new%20era%20of%20global%20flows/MGI-Digital-globalization-Full-report.ashx>

¹¹⁶⁹ 'Box 4. The impact of global flows on employment' in Manyika *et al.*, *ibid.*

1.3. Initiatives with young people will be important for the future, and will need to address some concerning deficits of participation in key subject areas. The recent review of post-16 mathematics in England by Professor Sir Adrian Smith (the Smith Review) was prompted by evidence and concern that the proportion of students who choose to continue to study mathematics after the age of 16 is much lower in England, Wales and Northern Ireland than it is in other comparable countries.¹¹⁷⁰

1.4. RSS's own view is that to meet the future needs of industry relating to data science, machine learning and AI, young people need not only have a strong mathematical and statistical education but also strong practical experience of digital and data analysis, and that all should get the chance to analyse real data using technology. Greater participation in this at all levels would support the use of data in the economy and in society, and would help to diversify employment in the science and technology sector. New 'core maths' qualifications for England are a development which we particularly support, as these should boost participation among students who would not otherwise continue to study mathematics post-16.

1.5. A wealth of other evidence for a digital skills gap has also been published, for example in the House of Commons Science and Technology Committee's report on the 'Digital Skills Crisis', which recommended that "the Government needs to establish an effective pipeline of individuals with specialist skills in data science, coding and a broader scientific workforce that is equipped with a firm grounding in mathematics, data analysis and computing."¹¹⁷¹ The Smith Review has further recommended that the Government should commission a study into the long-term implications of the rise of data science, to look at the skills that are required for the future.¹¹⁷² The RSS would be supportive of action on this, alongside actions that support quantitative education and access to jobs in the data economy.

2. The role of government

2.1. We welcome the commitment made to research and innovation in the UK's industrial strategy, which will add £4.7 billion in investment before 2020-2021. However, the sources and provision of research funding will undergo changes,

¹¹⁷⁰ Nuffield Foundation (2010) *Is the UK an outlier? An international comparison of upper secondary mathematics*

http://www.nuffieldfoundation.org/sites/default/files/files/Is%20the%20UK%20an%20Outlier_Nuffield%20Foundation_v_FINAL.pdf, cited p. 29 in *Report of Professor Sir Adrian Smith's review of post-16 mathematics* ['The Smith Review'] [PDF], July 2017
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/630488/AS_review_report.pdf

¹¹⁷¹ House of Commons Science and Technology Select Committee (2016) *Digital skills crisis* [PDF], <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/270/270.pdf>

¹¹⁷² The Smith Review, *ibid*, p. 14

following the formation of UK Research and Innovation (UKRI) and the UK's exit from the European Union. As these changes take place, the UK needs to shore up its position in AI and related disciplines: the overall number of graduates from mathematical science degrees (undergraduate and postgraduate) has declined, and mathematics, statistics and computation have been listed as 'vulnerable capabilities and skills' within the UK's bioscience and biomedical research base. We look to the new UKRI to prioritise research capabilities in data science and AI, and in the mathematical sciences that underpin them, on a cross-cutting basis across all disciplines. It could ensure that there is national support for this across the UK's research and innovation system, to an extent that individual research councils are unable to do.

2.2. National investment in AI and data science may benefit from review. Jo Johnson, the Minister of State for Universities, Science, Research and Innovation, has highlighted the UK's concentration of public investment in the 'golden triangle' (which refers to leading universities in London, Oxford and Cambridge).¹¹⁷³ The Alan Turing Institute has been established in London with a focus on data science. The RSS would be supportive in principle of expanding the scope of the Alan Turing Institute's activities to a much wider network of research and teaching institutes, to be a truly national supporter of data science and AI. In parallel with developing the UK's skills base, strong and sustainable international avenues for recruitment into research, teaching and training need to be maintained, to ensure that access to talent is not unnecessarily affected by Brexit.

2.3. The quality of data that is accessed and used for AI is easily obscured but forms a crucial basis for new developments. There is much that can be done to strengthen data infrastructure that supports new technology and innovation. We can access new sources of data for statistics on the economy and other domains, with data sharing across government and from the private sector enabled in the Digital Economy Act 2017. Greater access could be afforded to local data, and more could be done to release open data. There is also an enormous range of new *sources* of data coming into play, such as sensor technology and connected appliances, which could soon be widely applied in a variety of fields, including energy, transport, cities, and healthcare. The UK's data protection regulation – its adoption of the EU's General Data Protection Regulation and its subsequent approach within domestic legislation – will be crucial to AI and related fields.

¹¹⁷³ Department for Business, Innovation and Skills and Jo Johnson MP (2015) 'Speech: one nation science' [webpage] <https://www.gov.uk/government/speeches/one-nation-science> ; Higher Education Statistics Agency (2017) Income and expenditure by HE provider 2015/16 and 2014/15 (Table 1) [XLSX]. <https://www.hesa.ac.uk/data-and-analysis/providers/overviews?year=620&topic%5B%5D=606>

3. Ethics / Impact on society

3.1. Research that the RSS commissioned from Ipsos MORI in 2014 found a 'data trust deficit' among members of the public, who trusted organisations to a lesser extent on how they handle their data than they trust them generally.¹¹⁷⁴ The Government Office for Science has highlighted the range of potential benefits from adopting artificial intelligence (AI) and machine learning, which include making public services more efficient by anticipating demand and tailoring their provision, and making decisions more transparent.¹¹⁷⁵ However, to more fully realise these benefits, there is a need to address public concerns. Even those uses that are on balance well regarded by the public, such as the use of data for beneficial medical and public health research, can be badly affected by loss of trust.

3.2. Further research in this area points to a need for caution when widening the field of application for unexplained / partially explained data science and AI, from less regulated industries where they may have been developed, to those that require much greater explainability. We see important differences in the level of pressure to explain data science and statistical approaches across different industries. Developments in medicine and in clinical trials, for example, have become increasingly regulated to reduce potential harm, whereas other industries such as advertising, entertainment and online social media platforms are much more lightly regulated, and might remain so. The divergence between fields can lead to problems: Google DeepMind for example has said of its arrangement for data sharing from the Royal Free Hospital: *'we underestimated the complexity of the NHS and of the rules around patient data, as well as the potential fears about a well-known tech company working in health. We were almost exclusively focused on building tools that nurses and doctors wanted, and thought of our work as technology for clinicians rather than something that needed to be accountable to and shaped by patients, the public and the NHS as a whole. We got that wrong, and we need to do better.'*¹¹⁷⁶

3.3. With the growth of digital and data infrastructure there is growing scope for failure in how such data are accessed and used. For important societal applications (e.g. in the labour market, for access to jobs or for appraisal of performance) we believe there should be scope for appeal by members of the

¹¹⁷⁴ 'RSS research finds 'data trust deficit', with lessons for policymakers' [webpage], *StatsLife*, 22 July 2014. <https://www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers>

¹¹⁷⁵ Government Office for Science (2016) *Artificial intelligence: opportunities and implications for the future of decision making* [PDF] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf

¹¹⁷⁶ DeepMind (2017) 'The information commissioner, the Royal Free, and what we've learned' [webpage], 3 July 2017, <https://deepmind.com/blog/ico-royal-free/>

public who may be badly affected, as well as scope for the organisations that use such algorithms to evaluate the decisions that were taken and on what basis. Transparent and defensible statistical outputs should ideally be the end goal of innovation in these areas. In circumstances where this is not the case, developments should have a level of explainability in mind to avoid key failures for their industry and for service users. It is important for existing law (e.g. anti-discrimination) to develop through the courts to manage newly arising challenges.¹¹⁷⁷

3.4. The RSS has long suggested the establishment of a body to take forward thinking on data ethics in the UK. In autumn 2015, we held a workshop on the opportunities and ethics of big data, and suggested formation of a national council for data ethics to the Science and Technology committee, who then made it a recommendation in their Big Data Dilemma report.¹¹⁷⁸ The idea of a data ethics body has been gaining momentum ever since. The Conservative Party's 2017 election manifesto committed to setting up a Commission on the Use and Ethics of Data, and a report from the Royal Society and British Academy has recommended a new stewardship body for data governance.¹¹⁷⁹ The Nuffield Foundation, RSS, Alan Turing Institute, Royal Society, and British Academy are engaged in a partnership to take forward thinking in this area.¹¹⁸⁰

3.5. Deliberation is particularly important, as approaches to transparency and accountability will not be adequately addressed by legislation. Researchers from UCL Big Data Institute for example consider that transparency cannot be

¹¹⁷⁷ Royal Statistical Society (2017) 'The use of algorithms in decision making: RSS evidence to the House of Commons Science and Technology Select Committee Inquiry' [webpage].

<http://www.rss.org.uk/Images/PDF/influencing-change/2017/RSS%20evidence%20on%20the%20use%20of%20algorithms%20in%20decision%20making%20April%202017.pdf>

¹¹⁷⁸ The Royal Statistical Society (2016) *The Opportunities and Ethics of Big Data: Report of a consultation run by St George's House in November 2015 in association with The Royal Statistical Society, supported by the British Academy and SAGE* [PDF]

<http://www.rss.org.uk/Images/PDF/influencing-change/2016/rss-report-opp-and-ethics-of-big-data-feb-2016.pdf>

Recommendation 14 in 'The big data dilemma: Government Response to the Committee's Fourth Report of Session 2015–16' [webpage], 26 April 2016. Commons Select Committee > Science and Technology.

<https://www.publications.parliament.uk/pa/cm201516/cmselect/cmsctech/992/99204.htm>

¹¹⁷⁹ British Academy and Royal Society (2017) *Data Management and Use: Governance in the 21st Century* [PDF]

<http://www.britac.ac.uk/sites/default/files/Data%20management%20and%20use%20-%20Governance%20in%20the%2021st%20century.pdf>

¹¹⁸⁰ 'Convention on Data Ethics' at Nuffield Foundation [website] » News » 'Nuffield Foundation announces additional £20 million research funding, Fellowship programme, and major data ethics initiative in new five-year strategy' <http://www.nuffieldfoundation.org/news/nuffield-foundation-announces-additional-%C2%A320-million-research-funding-fellowship-programme-and->

guaranteed and that sometimes the power of machine learning models may mean that a lack of transparency is justified:

"Modern machine-learning algorithms are typically designed to excel in predictive accuracy using massive volumes of data. The availability of extremely large datasets, together with modern computational power, makes this approach quite practical. However, with prediction as the endpoint, such algorithms tend to assimilate the input data and construct complex models with convoluted and interacting components. [...] It thus becomes difficult to unpick specific strands of the decision-making process to understand precisely how a conclusion was reached. By contrast, traditional statistical algorithms are concerned with explanation as well as prediction, and tend to use clearly specified, often linear models, which are easier to scrutinise –although they are, on occasion, less powerful. In some cases, the impressive performance of ML algorithms can make the lack of transparency a reasonable trade-off, but this may not always be the case." (Olhede & Rodrigues, 2017¹¹⁸¹)

3.6. Stronger public engagement will be needed. In connection with their review of governance for data management and use, the British Academy and Royal Society reviewed public dialogues and engagement in the UK over the past ten years regarding the collection, sharing and use of personal data. Public awareness of new uses of data, such as machine learning, is found to be low, and they find that very few studies have investigated public attitudes to new and future uses of data: "While some studies have explored potential near-term applications of data technologies, none so far have looked into future worlds enabled by data [...] While several studies have looked into what criteria people use to define what is considered a valuable and beneficial output of data, they have not looked in depth at the social and ethical values at stake nor at the tensions between public good and personal risk."¹¹⁸²

3.6. New thinking on the application of ethical principles can also be driven forward by learned societies and professions. Data science is relatively young as a profession, with few professional standards. The application of ethics should be better understood, with strong ethical training embedded into data science courses, so that data scientists can anticipate issues including with the data that they train their algorithms on. Professional bodies should also take a leading role in developing standards, and the Data Science Section of the Royal Statistical Society is willing to help in this regard.

¹¹⁸¹ Olhede, S. Rodrigues, R. (2017) 'Fairness and transparency in the age of the algorithm', *Significance*, 5 April 2017. <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2017.01012.x/full>

¹¹⁸² p. 4 in British Academy & Royal Society (2017) *Data governance: public engagement review* [PDF], <https://royalsociety.org/~media/policy/projects/data-governance/data-governance-public-engagement-review.pdf>

12 September 2017

The RSA – Written evidence (AIC0157)

1. The pace of technological change

- 1.1. Artificial intelligence can be defined as computing software that completes tasks typically requiring human intelligence. More specifically, AI is an algorithm, or a bundle of algorithms working in unison, that follow a series of steps to arrive at an action or conclusion.
- 1.2. It is only in the last two decades that artificial intelligence has begun to live up to the hype set by its original founders in the 1940s and 50s. This is due to three key breakthroughs: (i) new approaches to building AI systems including machine learning and deep learning; (ii) a mammoth increase in the amount of data available to train these systems, owing to the advent of the internet and the subsequent spread of internet-enabled smartphones; and (iii) increasing computer power that has broadly followed Moore's Law.
- 1.3. The introduction of machine learning methods has been particularly transformative. Prior to this, software developers would painstakingly write lines of code, resulting in an 'expert system' built on a series of if-then rules to guide decision-making. Machine learning instead trains algorithms working backwards from existing data. Taking the example of image recognition, a machine learning approach might begin by feeding an algorithm with a series of labelled images – say of a house, car, balloon or bicycle – which the algorithm would then use to create a generalised rule to determine whether a future image fits these categories. Deep learning is a subdomain of machine learning, and uses so-called neural networks to spot more sophisticated patterns in data.
- 1.4. The excitement surrounding artificial intelligence is warranted. New AI systems can detect cancerous skin cells as accurately as a trained pathologist, detect fraudulent transactions in a matter of milliseconds, coordinate complex flows of goods on behalf of logistics firms, trade stocks and shares on financial markets, and write articles for company reporting. However, according to some technologists, general artificial intelligence – systems with equal or greater all-round intelligence to humans – remains a distant prospect, and narrow AI – systems that undertake specific tasks such as image recognition – continues to be limited.¹¹⁸³ As the technologist and entrepreneur Gary Walsh recently put it, "Although the field of AI is exploding with micro discoveries, progress towards the robustness and flexibility of human cognition remains elusive".

¹¹⁸³ See for example: <https://techcrunch.com/2016/12/05/deepmind-ceo-mustafa-suleyman-says-general-ai-is-still-a-long-way-off/>

- 1.5. Nonetheless, AI will become increasingly sophisticated and capable, owing to greater computer power, deeper pools of data and further research into programming methods. The huge amount of funding now flowing into AI research will lead to new breakthroughs. The US government alone invests \$1.1 billion in unclassified R&D around AI. One factor that might slow the development of AI is public resistance, particularly if its deployment leads to catastrophic outcomes such as a major cyber breach.

2. Impact on society

Artificial intelligence and work

- 2.1. Fears of mass job destruction at the hands of AI are likely to be overstated.¹¹⁸⁴ First, and as noted above, there are still many functions that are beyond the reach of machines, such as those relating to creativity, social intelligence and manual dexterity. Second, AI is more likely to automate tasks rather than whole jobs, allowing workers to pivot into new roles. And third, AI systems will complement what workers do, not just compete with them. A good example is the use of chatbots in customer service that can draft partially automated responses for staff to then edit and refine. Another example is medical software that can read through reams of articles and textbooks to help doctors diagnose conditions.
- 2.2. The emergence of AI will also directly give rise to new types of work, for example in creating, overseeing and maintaining machines. Our analysis of the Labour Force Survey shows that the number of programmers has grown by 40 percent since 2011, while the number of IT directors has more than doubled over the same period. While these jobs alone may not make up for the number of those lost to artificial intelligence, they will spawn extra jobs in ancillary service sectors that exist to serve their needs, whether in the entertainment, retail or healthcare sectors. Berkley economist Enrico Moretti suggests that every new job in the tech sector has the potential to generate 5 new jobs elsewhere.¹¹⁸⁵
- 2.3. Taken together, we believe that jobs are more likely to evolve than to be eliminated in the wake of AI's development. The question then becomes one of technology's impact on job quality rather than job quantity. An upcoming RSA report will urge economists and policymakers to pay closer attention to how AI will impact other aspects of work beyond job availability, including recruitment, pay, productivity, autonomy and the overall purpose and meaning attached to jobs. For example, algorithms used to screen candidates in workforce recruitment could exacerbate existing biases (if they are trained on biased decisions), or equally they

¹¹⁸⁴ See forthcoming report: Dellot, B. and Wallace-Stephens, F. (2017) *Age of Automation*. London: RSA.

¹¹⁸⁵ Moretti, E. (2010) 'Local Multipliers' in *American Economic Review*, Vol. 100, No. 2, pp. 373-77

could ensure that candidates are only selected according to their experience and qualifications.

- 2.4. Whether or not AI is a burden or a blessing to workers therefore depends on the choices that are made by employers, educators, policymakers, as well as the companies building the technology. However, our research suggests this debate could be somewhat redundant if the take-up of AI in the UK economy is as low as it currently appears. Our RSA/YouGov survey found that only 14 percent of business leaders have invested in AI and/or robotics, or are soon planning to. 20 percent want to invest but it will take several years before they can 'seriously' do so. A further 29 percent say the technology is either too costly or has not been properly tested. (N.b. While the respondents were asked to consider both AI and robotic technologies, it is likely the former is relevant to more businesses and sectors).
- 2.5. There may be some who believe the slow take-up of AI would be good news for the labour market, in the sense of sparing workers disruption and the possibility of losing their job. Yet this is a short-sighted view that ignores the parlous state of work for many people today. Average wages have yet to recover to their pre-crisis levels owing to the slowest decade of earnings growth in 150 years. This is a reflection of lacklustre productivity levels, with UK workers on average 35% less productive than their counterparts in Germany and 30% less productive than workers in the US.¹¹⁸⁶
- 2.6. Deployed and managed in the right way, we believe artificial intelligence can put the UK on the path to a better world of work. AI could raise productivity levels, generate the wealth necessary for pay growth, phase out dull, dangerous and dirty work, and allow more human-centric and intellectually stimulating jobs to prevail. Other studies have come to the same conclusion. The LSE's Leslie Wilcocks, who has studied the effects of automation in industries such as energy and journalism, concludes that in most cases "jobs were reconstructed, and expanded, rather than wholly automated... we found staff not feeling threatened by automation but instead appreciating having fewer repetitive tasks".

Artificial intelligence and society

- 2.7. As well as transforming how we work, artificial intelligence will also influence our lives as consumers, learners, voters, patients and citizens in the round. AI has the potential to improve living standards and tackle some of society's most stubborn challenges. For example:

¹¹⁸⁶ Be the Business (2017) *Top business leaders call on the UK to tackle productivity gap at the launch of 'Be the Business'* [Press notice] Available here:

https://www.bethebusiness.com/wp-content/uploads/2017/07/launch_of_be_the_business.pdf

- Healthcare diagnostics – Algorithms may extend lifespans by enabling faster and more accurate diagnostics. Using deep learning approaches, Kings College London was able to double the accuracy of brain age assessments using raw data from MRI scans. Similarly, DeepMind have been working with Royal Free Hospital to create an AI-powered device called Streams that can quickly review and screen test results for diseases such as acute kidney injury.
 - Drug discovery – US-based startup Recursion has combined automated microscopes (robotics) with image recognition software (AI) to rapidly test the impact of beta drugs on unhealthy cells, leading to the identification of 15 potential treatments. GlaxoSmithKline believe their investment in AI could cut drug development time down from an average of 5.5 years today to just one year.
 - Agricultural efficiency – Israeli tech company Prospera has developed a device that uses a combination of cameras and machine learning algorithms to detect early signs of crop disease, enabling farmers to step in early to save harvests. Elsewhere, a company called Blue River Technology has developed a roving robot imbued with AI that can meticulously pinpoint and eliminate weeds in crop fields, leading to higher yields.
- 2.8. Equally, artificial intelligence has the potential to cause significant harm. AI systems could compromise people’s privacy, undermine democratic elections, weaken media and accurate news reporting, power the use of autonomous weapons, deny people access to vital services such as insurance, and expose our institutions to more sophisticated cyberattacks. Potentially malevolent uses of AI have already been revealed:
- Biases in criminal justice – A number of US courtrooms use artificial intelligence to inform judgements on bail and custodial sentencing. An investigation by US media outlet Propublica found that an algorithm used in a Florida courtroom was twice as likely to falsely flag black defendants as future criminals as it was white defendants.¹¹⁸⁷
 - Monitoring in the workplace – Several start-ups are developing software to log staff behaviour on office computers, including browsing history, email messages, keystrokes and document use – data which is then used to create a baseline of employee performance and flag instances of underperformance.¹¹⁸⁸
 - Shaping voter behaviour – A journalistic investigation by the Observer in February 2017 revealed that AI may have been used to influence voter

¹¹⁸⁷ <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>

¹¹⁸⁸ Gilligan, A. (2017) *Bosses track you night and day with wearable gadgets* [article] The Times, 15th January 2017.

behaviour in the EU referendum.¹¹⁸⁹ The report claimed that machine learning algorithms were used to create targeted and 'highly individualised' adverts based on data harvested from Facebook.

2.9. Note that in many instances it is not clear whether an application of AI is acceptable or unacceptable according to social norms. With regards to elections, advertising (underpinned by focus groups and polling) has always been used to influence voter opinion. One might argue that deploying AI to shape voter behaviour is simply an extension of this activity. Likewise, in the case of workplace monitoring, it is not obvious where the dividing line sits between a use of AI that innocuously collects useful information and that which is overly intrusive.

2.10. A number of solutions have been mooted as a way of controlling and monitoring AI:

- **Ethical frameworks** – New standards for tech developers and companies to follow when creating AI
- **Software deposits** – Virtual access points where consumer rights organisations and government inspectors can audit commercial algorithms without compromising IP
- **Regulatory sandboxes** – Safe spaces for tech companies and tech adopters to experiment with new forms of AI, under the close supervision of regulators
- **Explainable AI** - The development of AI systems that can explain the steps taken to arrive at a decision, thus bringing a degree of transparency to the technology¹¹⁹⁰

3. Public perception

Understanding and engaging with AI

3.1. There is an argument for strengthening public engagement on the development and application of AI, for example through deliberative processes. These can foster an informed dialogue that engages with trade-offs, offers insight into ethical questions, takes into account citizen perspectives on opportunities, and highlights citizen concerns. Programmes such as the European Commission's Engage2020 (Engaging Society in Horizon2020) propose that public engagement of this kind is inclusive, anticipatory, reflexive (innovators are asked to consider more deeply their own standpoints and assumptions), as well as responsive.

3.2. More specifically, we believe the insights derived through public engagement initiatives can be used to:

¹¹⁸⁹ <https://www.theguardian.com/politics/2017/feb/26/us-billionaire-mercero-helped-back-brexit>

¹¹⁹⁰ It is currently difficult to decipher the decision pathways of some machine learning and deep learning algorithms – what is commonly described as a 'black box' problem. DARPA recently launched the Explainable AI (XAI) programme to further innovation in this space.

- Shape the behaviour of tech companies and developers so that innovation in the sector serves the wider public good
- Inform private investment decisions taken by venture capital firms, angel investors and other sources of capital
- Inform public investment decisions taken by institutions such as Innovate UK and university research councils
- Create an accountability mechanism by which to judge these investment decisions
- Inform the decisions of regulators about how to manage AI (e.g. in data protection, healthcare and finance)
- Strengthen the overall legitimacy of the technology and the companies that use it

3.3. The last of these is critical in preserving the social licence of AI and the companies that build and deploy it. Already there are indications that a large section of the public are hesitant about this technology. A recently commissioned opinion poll (2017) conducted by Ipsos Mori for the Royal Society found that almost one third of people believe the risks of ‘machine learning’ outweigh the benefits, while 36% believe the risks and benefits are balanced.¹¹⁹¹ Public engagement would help decision makers to understand what is and what is not an acceptable use of AI, and build consensus on when and how it should be used in a responsible manner.

3.4. Government departments have already prototyped models of public engagement on related, complex and controversial technology issues, such as robotics and the genetic modification of food, through a BEIS-funded 10 year programme called Sciencewise.¹¹⁹² The RSA’s own Citizens’ Economic Council has also brought together citizens with economists, policymakers and officials from organisations such as the Bank of England, local authorities, corporations and pension funds to engage in a similar dialogue about the goals of the UK economy.¹¹⁹³

3.5. Public engagement does not have to be limited to deliberation events that bring stakeholders together in person. MIT has built an online tool called Moral Machine to crowdsource public views on how and when AI should be used in specific contexts, such as with self-driving cars. We urge technology companies and research institutions to experiment with other innovative ways of collecting public opinion on AI.

3.6. As noted above, the use and application of artificial intelligence is already raising important questions about its implications for the quality of UK democracy. As social media platforms adopt machine learning in order to

¹¹⁹¹ Public Views of Machine Learning, Royal Society report (April 2017)

<https://royalsociety.org/~media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf>

¹¹⁹² Sciencewise: <http://www.sciencewise-erc.org.uk>

¹¹⁹³ RSA Citizens’ Economic Council: www.rsa.org.uk/citizenseconomy

hone and finesse messaging; and as they become particularly effective in targeting messaging at key democratic moments (for instance, General Elections and the EU referendum election), there is a need for dialogue about the most appropriate use of AI that contributes towards a better, more informed democratic debate; and that does not inadvertently contribute to a democratic deficit.

4. Industry

- 4.1. The RSA has been arguing since January 2016 that the UK needs to revisit competition law to account for new 'networked monopolies', which are created when platforms exploit the 'network effect' in order to scale. Typically, the network effect alone is not enough to sustain market power; rather, these companies also engage in strategies of empowering users to participate in shaping favourable regulatory practices that they too will benefit from as consumers. This exposes the weakness of anti-trust legislation, which is only enforced if consumers are being short-changed in terms of price or access to competitors rather than on account of their data being monopolised. The RSA has previously recommended that the competition act be modernised to safeguard a wider range of interests.
- 4.2. While some companies may arguably have data-based monopolies, there are other institutions that hold large swathes of data that are of interest to companies. Some concerns are being raised about how public institutions are sharing their valuable data with private companies. For example, it was recently found that the Royal Free NHS Foundation Trust in London failed to comply with data protection rules when it gave DeepMind, a British AI company acquired by Google, 1.6 million patient records for a trial. According to the Information Commissioner's Office (ICO), the deal breached UK data law because the Trust and DeepMind did not properly inform patients that their details were going to be used in the trial in the first instance, as well as how. An important consideration, however, is how this partnership was initially formed. DeepMind have since worked with patient experts to create a patient and public engagement strategy.
- 4.3. Guidelines should be produced to help public institutions in a position to share data for the wider good; for example, these institutions could create competitive tendering processes rather than forming partnerships behind closed doors. Public data can be as valuable as public money, particularly if it is being used to build a commercially viable product, so it should be subject to the same processes as public investment.

5. Ethics

5.1. Section 2 describes several ethical implications associated with the use of AI, including what it could mean for discrimination, privacy and fair and free elections. Other pressing issues include:

- The development of AI, and government investment in AI (what kind of AI-innovation should government take an active role in supporting and investing? Are there limits to the government's role in investing in AI and if so, what are they?)
- Issues concerning how effective the technology is, especially at predicting behaviour that might be unpredictable even by technology that is relatively advanced (i.e 'irrational' human behaviour, such as risky driving)
- Human bias and discrimination (perceived and actual) replicated by AI systems
- The risks taken by AI and who bears responsibility for that risk; as well as the implications for regulators, lawmakers and insurance companies in accounting for risk
- Situations in which AI makes choices to which there is no agreed moral consensus (for instance, in self-driving car scenarios where the technology may need to make a choice about harm)
- The extent to which AI is shaping the nature of discourse in our democratic society (through the proliferation/prevention of 'fake news' via social media, for instance)
- The implications of AI for the automation of work and the future of modern work
- Privacy and the right to individual liberty; as well as potential tensions with the need to ensure national security and the national public interest

5.2. These are questions that cannot be settled through the application of technical expertise alone – they require policymakers to gain a better understanding of the public's views on these issues in order to make decisions and create more effective governance about AI that retains democratic legitimacy and its social licence. These are also questions that play out in different contexts and in different scenarios where moral intuitions may pull in different directions.

6. The role of government and industry

6.1. The RSA therefore calls for an acceleration of AI and robotics, but in a way that delivers 'automation on our own terms'. Among our recommendations are to boost lifelong learning provision including through a new personal training account (like those launched in Singapore and France); to explore the merits of creating a UK Sovereign Wealth Fund that would invest in emerging technology and give every UK citizen a regular dividend; and to recalibrate our tax system so that the burden

of taxation falls more heavily on capital than on labour. We also urge employers to play a more active role in helping their workforce navigate future disruption, for example by co-creating automation strategies that set out how staff can work alongside new machines.

6 September 2017

Dr John Rumbold and Professor Barbara Pierscionek – Written evidence
(AIC0046)

**Dr John Rumbold and Professor Barbara Pierscionek –
Written evidence (AIC0046)**

[Submission to be found under Professor Barbara Pierscionek](#)

SafeToNet – Written evidence (AIC0087)

SafeToNet Artificial Intelligence Select Committee Response

6th September 2017

Richard Pursey, Chief Executive Officer, SafeToNet on behalf of SafeToNet

Introduction

1. As an award-winning, British technology company leading the way in safeguarding children online and protecting families by harnessing the power of artificial intelligence (AI), we have a unique view of the AI and wider technology ecosystem in the UK, and how the Government can best support its development. We therefore welcome the opportunity to contribute to the work of this extremely important Committee and timely inquiry.
2. Founded four and half years ago, SafeToNet's pioneering software uses machine-learning algorithms that can identify harmful content sent and received by children online, specifically on social networks, using their smart phones and other devices. Crucially, it can then block such content, in real time, before it is seen and the damage is done.
3. We are using AI to develop a cutting-edge solution to a growing problem – the need to keep children and young people safe online and protect them for challenges such as cyberbullying, grooming, sextortion and other predatory risks.
4. Our technology and the solution it offers, is far superior to the existing alternatives – faster and more effective. It exemplifies the huge potential of AI to solve and address key public policy challenges. Whilst we recognise the wider social and economic implications of advances in AI, we believe that if managed correctly it can deliver huge benefits for the UK.

The pace of technological change

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years?

5. At SafeToNet, and for the purposes of this submission, we would define AI as:

“the use of machine learning algorithms that allow machines and, specifically computer programs to think in the same way that humans think – with the ability to contextualise information and to use logic and intuition to make decisions.
6. To provide a practical example, our technology harnesses natural language processing and big data analytics to help it to detect and take the decision to block potentially harmful content. It does this based on its

understanding of the world, which allows it to assess the probability that something is harmful content.

7. Artificial intelligence is not a new concept, it has been around as an idea since as early as the 1940s. During the 1980s and 1990s, the initial promise of AI failed to be delivered. This period has been described by many as the “Winter of AI” when research and interest in it died down.
8. Resurgence in interest in AI is a relatively recent development with the technology giants recognising its potential and beginning to invest heavily in it. This was spurred on when a computer beat the world champion at the board game Go and the potential of AI to revolutionise the role that machines can play and the kind of tasks that they can undertake became clear.
9. Artificial Intelligence is not just about creating robots that can undertake tasks that human’s do, as it is often portrayed in the media. It has become a critical component of technology and tool in computer science which more and more programmes are harnessing. Cognitive computing which harnesses some form of AI, is already being used all around us, working side by side with humans to help us to manage complex issues and systems, to make decisions and to create new technologies. For example, city traffic management systems, credit check systems, Amazon’s SIRI, Uber and Tesla’s development of autonomous vehicles and Google’s search algorithms, all use some form of AI.
10. AI is also becoming more accessible and therefore used more and more. Increasingly you can purchase cognitive computing modules or systems that can be imbedded into your own technology, making it easier to harness AI – all you need is an idea and you can bring it to life without necessarily needing the in-house expertise. This gives SMEs and start-ups like SafeToNet readier access to such technology and will help to increase the volume and velocity of the development of AI.
11. As technology evolves, what we consider to be AI is also evolving. What was considered cutting edge AI ten years ago, such as speech recognition, is now seen as common place technology, with AI becoming synonymous with more and more advanced technologies.

What factors, technical or societal, will accelerate or hinder this development?

12. We are no doubt on the edge of a huge technological shift, which will see AI playing an ever-greater role in our everyday lives over the next twenty years. However, there are several factors that could hinder this development. These include:

- **Reticence amongst the user community to adopt the use of AI technology** – We already see reluctance amongst some sections of society to embrace new technologies such as internet banking because of fears over cybercrime, fraud and data protection concerns. This is often due to a lack of understanding and fear. There is also significant scaremongering in the media which feeds into fears about the development of AI promoting ideas that “the robots are coming”, as well as highlighting the negative impacts this could have such as eliminating low skilled jobs, rather than focusing on the benefits and the more positive applications.
- **Access to data** – AI systems are built on big data and require huge volumes of information to develop – the more data you put in and the more you teach the system the better the outputs – just like a human brain. Companies looking to harness AI, like SafeToNet, need significant volumes of data to inform their systems and ensure that they function effectively and achieve their full potential. For example, the more human behavioural patterns we put into our system, the better it will become at detecting threatening or offensive behaviour. As it stands small start-ups like ourselves can struggle to access data which is often in the hands of big multinationals like Google and Facebook or the public sector. We are keen to collaborate with such organisations. We are finding that there are considerable barriers to such data sharing – either due to reluctance on the part of the big companies to share data or due concerns over data protection. The risk is that the big data giants such as Google, Apple and Facebook become bigger and bigger based upon the data they are fed by small companies. They can therefore become more and more powerful if they resist sharing data.
- **Dominance of the big technology companies** – There is a clear risk that big technology corporations could become too dominant in the AI space, as they continue to invest heavily buying up the brightest talent and companies. As AI becomes more advanced it develops more and more quickly, meaning that early advances will give companies an immediate edge and they could then create barriers to entry for smaller firms. This could make it difficult for SMEs – often the most innovative companies – to get into the market hampering innovation.

Addressing these issues could help to accelerate the development of AI.

Is the current level of excitement which surrounds artificial intelligence warranted?

13. The current level of excitement which surrounds AI is indeed warranted. There is some fantastic technology being produced, with real potential to radically change our daily lives and address key challenges we face. We live in an age where we are only limited by our own imagination.
14. We are hearing more and more about advances in AI, but developments take time. The truth is that we don't yet know what this technology will yield and how far away we really are from a future where machines are more intelligent than humans.
15. The system is not perfect and humans will continue to need to work alongside these systems – at least in the short to medium term – to avoid false positives. However, the nature of machine learning means that the technology will then become more and more accurate the more data it processes developing a degree of autonomous decision making and being able to learn from experience, meaning that they will not make the same mistake again.
16. Whilst its potential is huge, the term AI can be overused. AI has become a buzzword for the technology sector – companies talking about and developing AI are able to attract significant levels of investment. As such it is our view that some overplay their hand and not all the technologies calling themselves AI warrant the name.
17. The broad definition of AI allows for this, but, there is a big difference between technology that harnesses some form of basic common-sense logic and what we would consider to be true "machine learning AI" – something which behaves like a human being – mimics human brain, takes in information and learns and can make decisions based on this information.

Industry

What are the key sectors that stand to benefit from the development and use of artificial intelligence?

18. AI can be applied to and provide solutions to a range of problems and public policy challenges across a range of sectors – be it the development of cleaner, cheaper autonomous vehicles which could revolutionise the transport sector or new medical technologies that can detect illness or support older people or those with disabilities to live independently. The education space also stands to gain substantially as AI systems better detect the intellect, intelligence and learning levels of a child adapting content and tutoring that suits the individual child rather than the mass demands of a volume-based curriculum.

- 19. At SafeToNet our expertise lies in harnessing AI technology to promote online safety and combat the online threats that children and young people face. We have therefore focused on the benefits that the development and use of AI can deliver in this sector in our answer, however many of the benefits will be common to other sectors.**
20. Every day we now scroll further online than we walk and with much of our lives taking place online this is increasingly where we are exposed to dangers that used to be confined to face-to-face interactions such as bullying, theft, abuse, grooming and sexual exploitation. Children and young people are particularly vulnerable to these threats and too often are being exposed to harmful content, which can have a lasting impact on them and parents, schools and others are struggling to protect them. Between 2015 and 2016 NSPCC figures suggest that there were 11,992 child sexual abuse images recorded in England alone - a figure that is up 64% from previous years.¹¹⁹⁴
- 21. Existing parental control systems aren't good enough because they don't block harmful content and prevent it from being sent and received and don't work across social networks. With the widespread use of mobile phones and other internet enabled devices, this is no longer enough. By harnessing AI, SafeToNet's technology offers a more effective, more nuanced solution, giving parents a better option than taking away children's devices and limiting their access to the internet. SafeToNet is therefore a social enabler – giving children the freedom to explore the internet and social web and to benefit from everything it has to offer safely.**
22. SafeToNet is a multifaceted piece of technology, which uses natural language processing (NLP), the ability for a computer environment to understand the words we're sending online – mixed with big data analytics (behavioural patterns). By that we mean behavioural analysis, sentiment analysis, contextual analysis etc. All of this goes into our software – a machine learning environment – our machine is learning from the mistakes it makes and becoming more accurate as time goes by just like a human brain would.
- 23. Essentially, SafeToNet uses AI to contextualise content allowing it to tell the difference between banter and aggression.** Then, once it learns more about the child it is protecting, it can accurately identify changes in their behaviour and threats that they might face such as cyberbullying, grooming, trolling, sextortion and other predatory risks. It can then address these risks by identifying harmful content or threatening behaviour and blocking either the content or the user.
- 24. SafeToNet is also more effective as it blocks content immediately – before it is seen and the damage is done. At present social media**

¹¹⁹⁴ <https://www.nspcc.org.uk/globalassets/documents/research-reports/how-safe-children-2017-report.pdf> last accessed 01.09.17

channels are moderated by humans – AI can do the same job much quicker and can process much larger amounts of data, meaning that it can deal with the sheer volume of negative and harmful content found online more effectively.

25. Using Facebook as an example, at present they have around 4,500 human moderators (with plans to hire an additional 3000 recently announced)¹¹⁹⁵, tasked with moderating the content on the site and dealing with reports of abusive or inappropriate content from its more than 2 billion monthly active users worldwide. It is estimated that 300 million photos are posted on Facebook every day, so even with 7,500 people reviewing just these photos, it would be an impossible task for them to take down every piece of harmful content.¹¹⁹⁶ Consequently, in March, a BBC investigation found Facebook failed to remove more than 80% of sexualised or abusive images of children¹¹⁹⁷. This is a huge failure and it's clear that human moderation is not working.
26. ***Human moderation has also proved to be distressing for the people undertaking it, whereas with SafeToNet humans never see the content it simply analysed by computers.***
27. In 2016 the Internet Watch Foundation found that 35% of child sexual abuse imagery takes humans more than 120 minutes to take down¹¹⁹⁸. ***SafeToNet's technology can block harmful content in less than a second make it much faster.***

How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and well-functioning economy?

28. Big technology giants like Google, IBM and Microsoft are pouring huge amounts of resources into the development of AI – buying up companies and the brightest and best minds and nurturing talent in this space¹¹⁹⁹. However, their monopoly over data, can and is, holding back innovation in this space. This is because SME's with limited resources are at a significant disadvantage in terms of accessing the data they need to develop their technologies further and to realise their full potential. Yet these small

¹¹⁹⁵ <https://www.theguardian.com/news/2017/may/21/facebook-moderators-quick-guide-job-challenges> last accessed 01.09.17

¹¹⁹⁶ <https://zephoria.com/top-15-valuable-facebook-statistics/> last accessed 01.09.17

¹¹⁹⁷ <http://www.bbc.co.uk/news/technology-39187929> last accessed 01.09.17

¹¹⁹⁸ https://annualreport.iwf.org.uk/assets/pdf/iwf_report_2016.pdf last accessed 31.08.17

¹¹⁹⁹ <https://www.economist.com/blogs/economist-explains/2016/04/economist-explains> last accessed 01.09.17

companies and tech start-ups, like SafeToNet are often leading the way in developing AI technologies for the public good.

29. As stated already in this submission, data is crucial for the development and advancement of AI and such companies have access to the 'big data' required to develop AI at scale and at pace, giving them a considerable upper hand. At present their focus is on developing products for commercial gain. Taking internet safety as an example, Facebook and Twitter have both come under fire for failing to address the spread of harmful content on their platforms. Both have the vast resources, technological capabilities and the data to replicate SafeToNet's technology and address this problem – in fact they could do it more easily – however they choose not to because it is not in their commercial interests at present.

30. We do not want to see only the big players engaged in development of AI, as this will not deliver its true potential. ***It is therefore important that SMEs can access big data and work collaboratively with these corporations, to ensure that their data can be harnessed for the public good. This is crucial to creating a vibrant, competitive market that truly fosters innovation and benefits the public.***

31. ***Government should play a more active role in incentivising and facilitating collaboration and data sharing between large corporations and SMEs. We also need a clear legal and regulatory framework in place, which recognises the need for data sharing and makes it easier and quicker to do so, whilst ensuring that personal data is managed safely and securely protecting individuals' rights and privacy.***

What role should the Government take in the development and use of Artificial Intelligence in the UK? Should Artificial Intelligence be regulated? If so, how?

32. With lots of debate about the development and future of AI and its potential, it is vital that the UK Government understands the opportunities and challenges this presents and can respond to them effectively – ensuring that we can fully harness its potential.

Providing a regulatory framework

33. As AI develops further and we get more and more sophisticated uses such as autonomous vehicles, regulation will be required to ensure that it is developed and used safely. However, it is vital that Government takes a measured approach to this and does not overregulate and stifle the development of new technologies. Equally, it is important that they start planning for this now and are properly prepared. We have already seen the challenges posed by the failure of regulation to keep pace with the rapid development and growth of the digital economy.

- 34. The Government therefore needs to take a more flexible approach to regulation – it must be leaner, more agile and more responsive to deal with the emergence of new technologies.**
- 35. Government must also be careful about treading the narrow line between freedom of speech and data protection and privacy and consider the issues that the development of AI presents. For example, when does 'manufactured' personal data created by an AI environment to predict personal behaviour patterns become personal data?**

Data sharing

36. There is a need to change how Government and regulators think about data protection, recognising the centrality of data to the development of AI.
37. At present, the data protection regime is centred purely on the rights of individuals and the duties of 'processors' towards them. This is important and of course of paramount concern, however we believe there should be different types of processors allowing for the recognition of those looking to 'process data for good'. For such processors, there is a need to find better ways to facilitate collaboration and make data sharing – in the right way, with the right safeguards in place – possible.
38. At SafeToNet, we recognise the significant responsibilities we have as a company processing huge amounts of personal data and we take these responsibilities extremely seriously. We believe this is crucial to build trust and understanding with customers – particularly as SafeToNet was designed to protect families from harm. We therefore prioritise the privacy of our users, whether that is parents, their children or people they are interacting with online. We ensure that we are fully aligned with the current data protection regulation and prepared for changes, such as the incoming EU General Data Protection Regulations (GDPR).
39. However, we need access to more data to improve our service and we are keen to work with the police, schools and other public-sector bodies to share data to develop our product further and to make sure it is best placed to keep children and young people safe. These organisations have huge amounts of behavioural data at their disposal – patterns which could help us to vastly improve our software's ability to detect and block threats.
40. As it stands this is extremely difficult, if not impossible. Such organisations recognise the value of our product, but they are hamstrung by existing data protection rules, which prevent this. This data could be anonymised and would be processed by our system and would not be seen by humans – indeed none of the content and data our system processes is ever seen by humans, safeguarding the privacy of the children we protect.
- 41.** We recognise that there is a delicate balance that must be struck here, but the current approach is limiting innovation, particularly for start-

ups like SafeToNet, who cannot otherwise gain access to such data. ***There is a clear need for renewed debate and discussion about data sharing and how we can build more flexibility into the system, whilst still protecting people's rights and privacy.***

Promoting good governance

42. Government also has a role to play in helping companies to navigate the ethical challenges associated with the development and use of AI.
43. It is crucial that companies and organisations harnessing AI recognise the potential ethical issues and take these concerns into account when using AI. At SafeToNet, we have an independent advisory board of experts that helps us to manage the ethical implications of our technology and we keep these under review.
- 44. *Good governance is important, but we also need a clear framework to operate within, which is where the Government has a role to play in terms of setting out minimum standards and furthering the debate about what is and what is not acceptable.***

Supporting British start-ups and innovative SMEs

45. There is potential for UK to be a leading player in the development of AI. Some fantastic technologies are already being developed in this country – SafeToNet is part of the Telefonica Wayra Accelerator Programme in London and we have seen first-hand some of the exciting and truly pioneering innovations being developed harnessing AI. The UK is becoming a real centre of innovation with a thriving digital economy.
- 46. *Government must recognise the potential and value of AI and how it can be applied to public policy challenges and find ways to support the development of technology for such uses, particularly when it is being developed by British SMEs, helping to ensuring that we remain at the forefront of such technology.***
- 47. *Specifically, in the internet safety space, Government must not rely on the large corporates to address these problems, but should help to empower and support SMEs like SafeToNet, who are already actively addressing the challenges of online safety and offering a clear solution.***
- 48. *Finally, Government must ensure that Brexit – negotiating the UK's departure from the EU and planning for the future does not overshadow everything else over the course of this Parliament. In the extremely fast-moving world of technology and digital progress, the UK must not get left behind in harnessing the power of AI and the benefits it can deliver safely and securely.***

5 September 2017

Sage – Written evidence (AIC0159)

What are the implications of artificial intelligence? Lords Select Committee on Artificial Intelligence Written Evidence by Sage – September 2017

Background

Sage, a UK FTSE 100 tech company is the market and technology leader for over three million businesses manage everything from money and people.

We strive to make business builders, from startups to scale-ups and enterprise more productive, tackling the admin that holds business owners back with the most intuitive and flexible cloud-enabled software.

Our mission is to help transform lives and create entrepreneurial opportunities for local communities around the world.

Sage – key facts:

- £1.6bn revenue including 70% recurring revenue from subscription software
- 3m customers across 23 markets
- Headquartered in Newcastle
- 53% employees paid by Sage software in UK
- Sage Foundation, our philanthropic initiative donates a unique 2% of time, cash flow and donated licenses
- Customer money flows £3trillion+ per year and almost £1tr in UK

Sage & AI:

- Target to deliver Invisible Accounting by 2020 delivered through AI/Machine Learning and collective intelligence
- Sage launched the first accounting chatbot Pegg, which is also gender neutral, in line with the first principles and codes of ethics for AI for business, published by Sage in June 2017
- Our Foundation is using AI to help combat Gender Based Violence in South Africa

Sage and AI

Sage's machine learning and AI journey began in 2016 when we launched Pegg, the world's first accounting automation personal assistant. Powered by AI, Pegg makes managing business finances as easy as texting a friend via popular messaging tools on Facebook and Slack.

Today we are embedding AI and Machine Learning into products across our portfolio, to help our customers cut the burden of administration, accelerate solving their problems and enhance the performance of their workforce.

New research published this month by Sage shows that companies currently spend an average of 120 working-days per year on administrative tasks.¹²⁰⁰ This accounts for around 5% of the total manpower for the average Small & Medium Sized Business, accounting is cited as the main burden. To put that in context, if UK businesses could be 5% more productive, this could lead to an increase in GDP of at least £33.9billion per year. Eliminating the admin burden on our businesses through automation could go some way to achieving this.

Sage introduced Pegg to help start-ups and small businesses execute routine **accounting** functions more efficiently. Designed to be mobile first, businesses can use Pegg to:

- ensure expenses are recorded
- determine the status of invoices
- generate outstanding invoice reminders
- check the overall balance of their business,

By texting requirements through a familiar style messaging platform on their phone, tablet or laptop, small business owners can run their companies and execute a large portion of their accounting administration quickly.

To help deal with the challenges of payroll administration, Sage extended the capability of Pegg to our **payroll** customer support services. Pegg allows users to ask questions about Payroll administration and compliance with HMRC requirements. Pegg improves the productivity of the users and improves our customers' accuracy.

And Sage is applying AI to support **HR** functions, making it easier, quicker and more efficient to recognize and record great work. Today's modern work environments often mean managers are remote from their employees or those employees are working in matrix for project teams. The AI solution can help with accurately obtaining and capturing feedback about employees from colleagues and co-workers.

Behind the scenes, we are using machine learning and AI to prompt businesses with **automated insights**, for example benchmarking how they are performing and improving marketing spend. The information generated using Machine Learning (ML) and AI, will be extremely valuable for the users, providing previously unrecognized insights about business performance or even future challenges.

¹²⁰⁰ Research of 3000 businesses in 11 countries, including UK, for Sage by Plum / FTI Consulting Sept 2017

These use cases highlight some of the benefits of AI by enhancing productivity, improving accuracy and reducing customer wait times and frustration.

Summary

AI - the opportunity for businesses

For the millions of small and medium size businesses we serve, AI heralds the opportunity to improve productivity understand, learn and carry out key business tasks and automate processes so our customers can focus their sales, delighting customers and growing their business.

There's no doubt AI, in conjunction with emerging technologies like big data is poised to revolutionize our lives and make societal gains as we move ahead to the Fourth Industrial Revolution. However, with the rhetoric in the media being largely negative, and the acceleration of AI gathering pace, the wider concern is that the tech industry will lose focus on the implementing the guiderails as the apocalyptic scenarios steal the headlines.

We cannot take it for granted that the potential of AI to do good will be realised without more positive focus and strategy coupled with agile frameworks in place to tackle ethical principles of AI.

In our view the biggest threat we face right now is not an exponential risk to humanity or widespread job loss that we read about daily in the media, it is slow pace to 'Digital Britain' that means only 56% of UK businesses are adopting technologies which is restricting workforce participation in AI application.

And even though use of AI is becoming more widespread 46% of consumers in the U.S. and 43% in the UK admitted they have "no idea what AI is all about"¹²⁰¹

Societal influencers, including Government need to talk up the opportunities of AI, and reprioritise their approach to threats.

We need a clear strategy that will tackle the immediate priorities:

- Evaluating the social and business benefits AI offers
- Educating businesses and consumers about AI, demonstrating when they are using it, how it can make a positive impact and how they can protect themselves
- Scrutinising data ownership to ensure a level playing field
- Examining today's AI development to ensure it is based on ethical principles such as the principles Sage has published

¹²⁰¹ Sage used Google Surveys between August 2-5, 2017 to collect responses from 500 consumers in the United States, 500 consumers in the United Kingdom

- Broadening the skills base to tap into the broader skills than pure tech ones needed for future application of AI

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Current state of artificial intelligence

The concept of AI is not new. It has been around for over 60 years in various forms. It is foolish to think we can stem the tide of progress – the fact is that already, millions of consumers and businesses are using AI whether they realise it or not in their day to day lives.

Google search prompts, autocorrect, Facebook friend suggestions and predictive texts in SMS are all based on AI technology.

The rise in levels of data, driven by the amount of information we all store on line in tools like Facebook, LinkedIn, Twitter, Wikipedia, etc, has opened many more possibilities and led to a period of extraordinary innovation and simplification of the process of using AI over the last 3 years.

Google's TensorFlow, Amazon AI services, IBM Watson, Microsoft Cognitive services are making AI application possible, and accuracy rates are vastly improved compared with 10 years ago. Amazon's Alexa Technology has 95% voice recognition accuracy in understanding human languages and Microsoft report only a 4% error rate of language understanding.

AI functions well currently when it focuses on addressing a specific problem or function, automating repetitive tasks such managing people's finances, checking compliance, and helping with questions that are rules based. More advanced examples are self-driving cars, medical triage and supporting diagnoses. In truth, where AI has 'gone wrong', it is most often in cases of 'General AI' with no clear use case.

The next 5 years and beyond

Applications of AI will become far more widespread and the technology behind it will accelerate development. AI is different from any other technology revolution because it learns to think from neural networks and can learn to build its own code. Continuously becoming more intelligent, progress and uptake of AI will occur at an exponential rate.

The possibility exists for us to use AI to start to address problems at scale. A radiologist diagnosing cancer using machine learning will become a common scenario. It is going to become much more commonplace for AI to make decisions for us or to be used in relatively new areas like the creative industries in art or music.

Decision makers including business and Government need to talk up the opportunities of AI, and reprioritise their approach to threats that could hinder development.

- Evaluating the social and business benefits AI offers
- Educating businesses and consumers about AI, demonstrating when they are using it, how it can make a positive impact and how they can protect themselves
- Scrutinising data ownership to ensure a level playing field
- Examining today's AI development to ensure it is based on ethical principles such as the principles Sage has published
- Broadening the skills base to tap into the broader skills than pure tech ones needed for future application of AI – furthering the UK's current leadership in this industry

That way we move to a longer-term vision of trusted widespread application of machine learning and AI that is delivering a net benefit to society.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

The current enthusiasm about the application of AI is absolutely warranted given the opportunity to solve problems facing us today, in fact we think the benefits are underrated and should be more widely talked about. We welcomed the Accenture report '[Fuel for Growth](#)' revealing that AI could double annual economic growth rates by 2035 and improve labour productivity by 40%, but what about the wider social impact?

The reality is AI can perform a vital function where people are lacking. For example, giving essential expertise and intelligence to a small business that would not otherwise be able to afford a CFO or CMO, addressing staff shortages in the health sector or lack of resources to personalise learning in the education sector. We can develop applications to complement human thinking and address important gaps.

And, there's a bigger story to tell here about how AI can deliver social good and help address critical resource challenges in developing and first world countries in areas such as health, social care and education.

The Government could take a lead in understanding this opportunity for AI to improve quality and efficiency of clinical processes or learning opportunities and making a more personalised experience possible.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

In the future, we may need to look at radical policy measures such as taxing robots or introducing the right to a Universal Basic Income but right now corporations and governments need to develop the basic guiderails that will deliver a stable and secure foundation of AI for us to build upon.

In turn, we believe this will help to eliminate negative sentiment, and support people and businesses in adopting this new technology within an ethical framework, to improve productivity and address social need.

- A. Protecting consumers and businesses.** Consumers and businesses need to be prepared and able to use AI applications confidently, knowing their privacy is protected and their needs represented.
- B. Reframing the AI skills strategy.** The education and business sector need to focus on reframing our AI skills strategy to enthuse more people about the opportunities given the much wider skills set that will be needed to apply AI in the future – beyond coding.

A. Protecting consumers and businesses

Areas that need addressing include diversity, accountability, equality and social impact. For example, underlying data being built into AI applications must be from diverse sources or AI systems will have bias built into them. So for example, if you build a system that learns from Wikipedia – only 17% profiles of notable people are women – this bias will then be perpetuated in the machine.

At this stage, it is too early to say whether additional regulation, on top of current data privacy and internet security regulations will be needed, and this could risk stifling a technology that's in its relative infancy.

But we cannot ignore the threats if AI is not developed with principles in mind. In our view industry and Government need to work together to take initiative in an agile way.

In June this year Sage took a first step and launched 'Core Principles when designing AI for Business' calling on other companies to do the same.

Our 5 principles to keep corporate AI accountable have been published globally to raise awareness and provide reassurance to our customers.

CORE PRINCIPLE 1: AI should reflect the diversity of the users it serves.

We need to create innately diverse AI. As an industry and community we must develop effective mechanisms to filter our bias as well any negative sentiment in the data that AI learns from, and ensure AI does not perpetuate stereotypes. Unless we build AI using diverse teams, data sets and design, we are at risk of repeating the inequality of previous revolutions.

CORE PRINCIPLE 2: AI must be held to account – and so must users

We have learnt that users build a relationship with AI and start to trust it after just a few meaningful interactions. With trust, comes responsibility and AI needs to be held accountable for its actions and decisions, just like humans.

We disagree with the notion that technology is allowed to become too clever to be accountable.

We don't accept this kind of behaviour from other 'expert' professions, so why should technology be the exception.

CORE PRINCIPLE 3: AI should level the playing field

AI provides new opportunities to democratize access to technology especially because of its ability to scale. Voice technology and social robots provide newly accessible solutions, specifically to people disadvantaged by sight problems, dyslexia and limited mobility.

Our business technology community needs to accelerate the development of these technologies to level the playing field and broaden the talent pool we have available to us both in the accounting and technology professions.

CORE PRINCIPLE 4: AI will replace, but it must also create

The best use case for AI is automation – customer support, workflows, rules-based processes are the perfect scenarios where AI comes into its own. AI learns faster than humans and is very good at repetitive, mundane tasks and in the long term, is cheaper than humans.

There will be new opportunities created by the robotification of tasks, and we need to train humans for these prospects – allowing people to focus on what they are good at, building relationships and caring for customers. Never forgetting the need for human empathy in core professions like law enforcement, nursing, caring and complex decision making.

CORE PRINCIPLE 5: Reward AI for 'showing its workings'

As with training a pet, you reward AI for the behaviour you expect from it. Any AI system learning from bad examples could end up becoming socially inappropriate – we have to remember that AI has no idea what it is saying. Only broad listening and deep learning from diverse data sets will solve for this.

Whilst designing the reward mechanism for AI, we need to build in the reinforcement learning measures that teach the machine how it should be achieved, not just optimize the end-result.

In summary there is a role for governments to bring stakeholders together to agree common principles for AI adoption and the UK Government could take a lead in doing this as part of its Digital Charter initiative.

B. Reframing the AI skills strategy

Today the creators of AI are tech specialists, but increasingly in the future we will see widespread demand for people with softer skills that can train AI. This opens up possibilities to develop much more diverse cohorts of AI employees than focusing on computer science-based skills, as important as they are.

Whilst the Government's AI Review is set to increase funding for more traditional education routes like PhDs and Masters in AI computing, the reality is that the kinds of skills needed are going to be much more diverse and grass roots - including interaction, linguistics and creative arts skills.

A much bigger conversation and more diverse, less 'elitist' talent and skills strategy is needed encouraging both young people and adults with ability and an interest in these wider skills sets to think about a career in tech and getting involved in AI development and application even if they do not follow the traditional academic route.

This can be done through apprenticeship programmes, a broader 'tech curriculum' in schools, vocational higher education course as well as university degrees to attract a wider, more diverse pool of talent into the industry.

That is why at Sage our Foundation is setting up an AI skills training programme which we call our 'Bot Camp' to inspire young people in schools and bring forward a generation of AI 'trainers' we referred to earlier. We will be focusing on opening up new opportunities for young people living in more deprived areas, to think about working on AI applications as a future career possibility. Our investment is AI are tech specialists, but increasingly in the future we will see widespread demand for people with softer skills that can train AI.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

As outlined above the greatest opportunities are for creating more diverse tech roles, improving productivity and applying AI for social good. Conversely the risks we face are exaggerating negative consequences of AI, not adopting ethical principles to underpin AI development and failure to tackle the digital divide – where millions of consumers and businesses will not access the benefits AI can offer.

Business and the wider economy stand to gain from development and use of AI, but the digital divide risks leaving many behind

The UK's small and medium sized businesses and the wider economy stand to secure huge benefits from artificial intelligence which will make running a business intuitive and smart and increase productivity.

2. Machine learning will bring about a new era of invisible accounting and the more digitised business processes are, the better it performs. 25% of tech-savvy small businesses reported growth of more than 10% compared to last year. This compares to just 11% of the small businesses that had not adopted digital technologies experiencing 10%+ growth.¹²⁰²

Britain's businesses are relatively unproductive compared to their European counterparts. Despite the rapid growth, relatively high wage jobs, productivity tools and Gross Value Added tech and digital are bringing to our economy, too many UK businesses are missing out.

According to Be the Business for every hour spent at work we make 20% less than a company in Germany. SMEs are vital to the UK economy, and an increase in productivity of just 5% could lead to an increase in GVA of at least GBP 33.9 billion.

New research published this month by Sage shows that companies currently spend an average of 120 working-days per year on administrative tasks. This accounts for around 5% of the total manpower for the average Small & Medium Sized Business, accounting is cited as the main burden. To put that in context, if UK businesses could be 5% more productive, this could lead to an increase in GDP of at least £33.9billion per year. Eliminating the admin burden on our businesses through automation could go some way to achieving this.

¹²⁰² 2017. Based on 1,053 interviews by IDC with small business owners < 10 employees across the US, Canada, the United Kingdom, France, Germany, and Spain.

Yet with 45% of UK businesses saying they do not have a software solution to help take care of this, it is no surprise that they're spending so much time on accounting-related admin.

In summary, late adopters and the digitally excluded risk losing the most, and without a clear AI strategy for social good the huge potential benefits of AI will not be felt across society at large.

The failure of many early stage businesses to adopt digital processes must be a key priority for the Industrial, Digital and AI strategies being developed. The Government's Making Tax Digital programme, where even the smallest businesses would be required to use free or low cost digital software to submit their income and expenses quarterly would have led to millions of UK businesses adopting intuitive software, powered by AI, to understand their tax liability.

In our view Government's u-turn that led to the re-design of this programme in July has setback progress to a more 'Digital' Britain, which would have seen further application of AI within compliance software, cutting the admin burden for millions of businesses.

The potential societal benefits and 'AI for good' are so far underestimated.

AI applications to support better services in healthcare or education should be embraced by the public sector. Use of AI in the care sector is in action in Japan with 'carebots' but we need to tell the story in a different way. Assisted Living not only addresses a shortage of caregivers, it helps people to live on their own and independently and not go to a care home. Likewise, AI will fundamentally improve diagnosis, education programmes and scale-up opportunities for more people to access vital information or help.

The Sage Foundation is applying AI technologies to combat Gender Based Violence (GBV) in South Africa with "The Magic App". South Africa's femicide rate (murder of a women by their partner, husband, boyfriend or ex-boyfriend) is the highest in the world at almost 5 times higher than anywhere else. Most women only report an incident of abuse after the 10th – 15th time of it happening. However, to effectively prosecute the abuser, a detailed history of the abuse is often required.

In November, the Sage Foundation will launch a hidden 'Magic App' to support women affected by GBV by using their smartphone to record and document abusive incidents. Confidentiality and discretion will be of the most critical importance before going live with this App. A Facebook messenger based artificial intelligence assistant will provide information, advice, and location of the

nearest police station or shelter-type centre – the AI will learn all of this from publicly available information.

This is one example of how AI can support a hard to reach population where resources are scarce.

5. Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

AI is already creeping into our daily lives, digital personal assistants like Apple’s Siri or Amazon’s Alexa, chatbots like Sage’s Pegg solution, or Netflix’s recommendations, are all using AI. It will permeate through society as people become more reliant on the solutions, as trust is established and the value experienced.

Promoting Sage’s “Core Principles for design” to help establish a robust framework which, in turn can be used to create trusted solutions and demonstrate value, will be an important step in the awareness and adoption cycle.

In the near future, there will be a need for formal education in schools and universities to include introductions to AI and ML solutions for many students, not just those in computer science classes. Academia and researchers will rapidly adopt and use AI and ML to help with research projects and to solve the problems facing society using data. These approaches will be shared with students, so a fundamental understanding of how to manipulate and interrogate data using AI solutions, will become critical components of any student’s capability.

More broadly, as outlined above, the Government has a key role to play in improving our understanding of the positive social impact that AI technologies have potential to deliver. This could start with an ‘AI for public services’ strategy.

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not? In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

Undoubtedly some sectors are more labour intensive and ripe for automation. However, this does not necessarily equate to net job losses. At Sage, we work with 14,000 accountants, supporting them with data driven, cloud-based solutions that make processes more efficient and open up opportunities for more business intelligence and a higher quality service to clients.

Automation opens up opportunities for accountants to redefine the practice and what it means for their clients – on their own terms. Some accountants are already embracing these opportunities. Research shows that 86% see automation as creating more value for their clients, saying they would be happy for technology to make the administrative elements of their job invisible, so they can focus on their clients and building their business.¹²⁰³

On the flip side 38% of accountants see emerging technology as the biggest threat to the accountancy profession and a third are still using manual methods for record keeping. By not adapting to the way the industry is changing, accountants are leaving their practice vulnerable to disappointing their clients and putting themselves behind the competition.

7. How can the data-based monopolies of some large corporations, and the 'winnertakes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

At Sage, we are focused on supporting and empowering small and medium sized business to be as successful as possible and competitive in their chosen field.

Historically, access to new and emerging technologies has been the luxury of large business, with high costs being a significant barrier to adoption for small and medium sized companies.

As the Internet of Things (IoT) and the hyper-connected society grow, machine learning and AI solutions will be needed to process and manipulate the vast quantities of data being generated. Access to massive volumes of data will not be limited to large corporations, companies of every size, even individuals, will be generating almost unfathomable volumes of data points. Processing and interpreting the data will need AI, whereas understanding the implications or ramifications of the output will require specific human skills.

The introduction of cloud computing technologies has created a sort of disintermediation, allowing SMBs access to data, access new cloud services to support their business, resulting in reduced capital expenditures, in such a way they can compete with larger corporations.

Data is often referred to as the 'new oil' with huge wealth potential for the companies rushing to create, capture and mine insights about their business, operations and customers. Many industries such as mobile telcos have invested heavily in data warehousing and analytic capabilities but it is only with the advent of ML and AI that they are truly able to capitalize on the information.

¹²⁰³ 'The Practice of Now' Sage

<http://www.sage.co.uk/lp/~media/918D19B8A1BA4DDDBD3E8EAB38627030.ashx>

One of the biggest challenges for many companies and organizations now will be monetizing the value associated with the data they and their business creates. Some data will be competitively sensitive, with companies have no desire to release or sell the accumulated data, whereas other data sets will perhaps be irrelevant or redundant to one organizations.

So in summary we see the introduction of cloud based solutions with the addition of ML and AI as key contributors to the democratization of new technologies, which should increase access to data for many more SMEs.

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

See Q3

9. In what situations is a relative lack of transparency in artificial intelligence systems (so called 'black boxing') acceptable? When should it not be permissible?

Many algorithms are commercially sensitive so it is unrealistic to expect companies to develop them publicly but that should not get in the way of improving transparency.

But, in our view, transparency will be key to improving trust and confidence in AI systems. As we set out in our Principles, it is in everyone's interest, including businesses, that AI should be rewarded for 'showing its workings'. It is important we should not get to the point where, as society, we blindly accept the information or conclusion that is derived from an AI system without the ability to question the output.

To maintain the desired transparency, it should be possible to complete many of the functions manually by following the required process or reverse engineering the solution. For auditing requirements, this could be done through a random selection analysing outputs relative to inputs.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

See above. We think Government has a critical role to play in improving understanding of the wider social and economic benefits AI has to offer, to encourage best practice and an ethical approach and to inspire a more diverse workforce and AI development.

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

No comment.

6 September 2017

Professor Maggi Savin-Baden and Eur. Ing. David Burden – Written evidence
(AIC0061)

**Professor Maggi Savin-Baden and Eur. Ing. David Burden
– Written evidence (AIC0061)**

[Submission to be found under Eur. Ing. David Burden](#)

SCAMPI Research Consortium, City, University of London – Written evidence (AIC0060)

Response to Lords Select Committee on Artificial Intelligence: Artificial Intelligence to Improve the UK's Health and Social Care

Rt Hon Paul Burstow, Professor Neil Maiden, Dr Dympna O'Sullivan, Dr Simone Stumpf, Members of the SCAMPI Research Consortium, City, University of London

SCAMPI, Artificial Intelligence and Health/Social Care in the UK

The EPSRC-funded Self-Care Advice, Monitoring, Planning and Intervention (SCAMPI) Consortium is researching new artificial intelligence technologies to support people with chronic diseases to improve the quality of their lives at home. Until 2020, it will develop and evaluate automated planning, reasoning and sensing technologies to support people with two conditions – dementia and Parkinson's disease – to enable them to plan and monitor their lives and care at home. These technologies will then be evolved and rolled out to support people with other chronic health conditions. It is hoped that SCAMPI will have significant future impact on people with chronic diseases and their families by using artificial intelligence in everyday health management decisions, and on third-sector organisations seeking to leverage these new technologies to solve critical health and social care challenges.

SCAMPI's new uses of artificial intelligence to support social care are important. The project is one of the few in the UK to deploy artificial intelligence in social care to empower people with new knowledge and capabilities, rather than to automate and replace these people. To respond to the Select Committee, SCAMPI draws on its knowledge and expertise from the perspectives of social care and related healthcare research. Key SCAMPI recommendations are:

1. Enable and educate the general public to take ownership of their personal health and social care data, as part of their active care and life planning;
2. Ensure that health and social care professionals are equipped to understand, procure and deploy artificial intelligence and machine learning through suitable informatics education and training;
3. To reduce the potential for incorrect decisions, increase the transparency of artificial intelligence algorithms to enable public scrutiny and oversight and intervention by health and care professionals;
4. Determine the mix of regulatory and procurement action necessary to ensure that black-box artificial intelligence does not deny people access to information generated from their own datasets – a risk to the ethical ownership of people's data;

5. Work with social care commissioners and providers to create opportunities for UK-based artificial intelligence research enterprises to support the sector realise the potential of these technologies; and
6. Regulators need to future proof the way they regulate. The changing landscape needs to be mapped against the scope Parliament has determined for each relevant regulator.

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

The state of artificial intelligence research and development is more advanced than most people in the general public believe, and in directions that most members of the general public do not recognise. Indeed, older artificial intelligence technologies in use are often not treated as such, for example the use of fuzzy logic in implantable devices such as pace makers. There are more established artificial intelligence developments in healthcare than in social care, and include opportunities to analyse big data quickly and reliably:

- Case-based reasoning and pattern recognition technologies that emulate how clinicians and other healthcare professionals reason, for example for diagnosis, decision support and care scheduling, from oncology to medical imaging. Machines pattern-match very effectively, leading to new potential for these algorithms to be applied to patient records more broadly, for example for screening for diseases by Deep Mind Health [1];
- Image analysis technologies, for example for spotting tumours and identifying skin cancer [2] and choosing successful IVF embryos [3];
- Natural language processing technologies applied to, for example, mining medical literature to provide decision support to treat cancers and discover new medical drugs [12] and medical chat-bots, as being trialled with the NHS [8];
- Multivariate analysis, which allows for contextual decision-making that is critical to people delivering good healthcare and social care. For example, NICE have used multinomial logistic regression to perform health technology assessment and is the decision making process that governs funding for health care systems [9].

There are limited applications of robotics as examples of autonomous artificial intelligence, for example to deliver mechanised companionship to older people with the *Paro* seal robot.

Developments to support social care are more modest, but include:

- The SCAMPI consortium's case-based technologies support creative thinking to manage challenging behaviours exhibited by people with dementia [4];

- Contextual and shared decision-making that is critical to delivering good social care.

SCAMPI predicts two factors to accelerate this development in the next 5-15 years:

- Synthesis of diverse technologies and data sources from different research disciplines is now possible. Moreover, the emergences of the Internet of Things and social media has enabled the simple, effective and scalable integration of different artificial intelligence technologies and data sources to support diverse health care and social care tasks;
- The strategic shift in healthcare from diagnosis and treatment, to wellness and prevention and self-care. For example, patient-centred applications already use the Internet of things, telemedicine and personalized care with available consumer devices such as the Fitbit and iPhone. Most perform trend analysis using captured data that can be inputs to existing and new artificial intelligence mechanisms. Example applications include Cellscope's digital otoscope [5] and AliveCor Kardia [6] - both add sensors to smartphones so that consumers can monitor ear infections or atrial fibrillation respectively.

To conclude, SCAMPI predicts new applications of artificial intelligence that integrate data from multiple sources to support people to deliver and to receive more personalised healthcare and social care. The focus will be to support, collaborate about and learn, rather than to automate care tasks.

2. Is the current level of excitement that surrounds artificial intelligence warranted?

Yes, it is warranted, but much of the research and development is misunderstood by most who are not familiar with the technologies. Most public debate is oversimplified, and focuses too much on robotics and on full automation of tasks that are currently undertaken by people at the exclusion of technologies that seek to enhance essential human skills. For example, in the social care sector, the scope for robots and the automation of carer activities are limited, let alone cost-effective, for most UK citizens and care services.

Instead, SCAMPI argues that, instead, greater benefits can accrue from focusing on technologies to enhance human knowledge and capabilities – technologies and people cooperating – rather than on automation. It also places a premium on the relational aspects of care and support and the quality of the human interaction. This might create fewer news headlines, but is a key direction of travel. Current examples of such technologies in healthcare and social care include the remote monitoring of people's health by healthcare professionals such as the TIHM project in Surrey [10], the use of telemedicine technologies, and personalised reminiscence therapy apps for people with dementia.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

There are different ways in which the general public can be prepared:

- Enhance awareness of what artificial intelligence is and how it is already been used, under people's control, in everyday activities. Once these technologies are established and accepted by most people, they are rarely considered to be artificial intelligence per se. Government has a role in educating the public as the presence, nature and uses of these artificial intelligence technologies – uses which are overwhelmingly benign rather than dangerous;
- Enable and educate the general public to take ownership of their personal health and social care data, as part of their active care and life planning;
- Ensure that healthcare and social care professionals who will interact with artificial intelligence systems are up-skilled to understand and exploit these technologies e.g. through health informatics education and training.
- Increase the transparency of artificial intelligence algorithms to enable public scrutiny and professionals to intervene, to reduce algorithm bias and the potential for incorrect decisions;
- Enable a deeper understanding of the ethical implications of using artificial intelligence in healthcare.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

SCAMPI argues that the UK public sector, and in particular UK social care, is currently benefiting little from the development and use of artificial intelligence, as few initiatives have been funded or reported. Instead, most current artificial intelligence serves business interests, particularly in industries such as automotive manufacturing, which can adopt automation quickly, and the needs of consumers with purchasing power, for example through easier access to information, improved decision making for purchases and greater convenience, as demonstrated by Amazon's online selling algorithms.

Artificial intelligence also rarely serves socially disadvantaged people and groups, for several possible reasons. Socially disadvantaged groups tend to lack the access to basic technologies needed to access artificial intelligence. There is little financial incentive for technology companies to invest in this sector, due to both the limited financial returns that are available and the funding crisis – a crisis that means that the sector is changing and evolving [11].

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Efforts should be made. Most members of the generic public know about artificial intelligence from news media, social networks and television and movies. Therefore:

- These different media need to present factual information about artificial intelligence, rather than alarmist stories about robots, automation and mass unemployment;
- Seek to educate the general public that most artificial intelligence will be developed to support and cooperate with, rather than automate and replace people, as countless examples of such technologies in healthcare demonstrate. Explain how hand-over or mixed-initiative control can already be coordinated between human and artificial intelligence, to show how people can take charge of these technologies when needed;
- Demystify with new white-box demonstrators that describe and explain exactly how systems operate in key domains such as education and healthcare, and where the intelligence is derived from;
- Direct software developers to build in more explanations of their artificial intelligence products and their outcomes, to encourage greater understanding, and encourage the public to use then create and customise their own artificial intelligence technologies, especially if it concerns their own healthcare and social care information.

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

The focus of SCAMPI is on social care, and to support and empower people, their families and their carers, rather than to automate complex care tasks and replace human carers. It is a key sector that stands to benefit. Alas, this is not a priority in artificial intelligence research and applications, due to relative lack of return on investment, and hence funding in the sector. The artificial intelligence providers have little real understanding of healthcare and social care challenges. And it remains difficult for traditional sectors like health to recruit the expertise needed to develop artificial intelligence solutions, due to the cross-discipline knowledge and expertise required.

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

Careful note has to be taken to reduce bias in artificial intelligence, to avoid simply replicating current decision-making and data collection biases. Existing biases in healthcare, and to a lesser extent social care data, exist because most data has been collected from white males, and which skews the analysis of data is skewed to that population, for example [7]. As a consequence, artificial

intelligence technologies risk deepening the emerging digital divide. Disparities can be mitigated through:

- Raising awareness in artificial intelligence researcher companies to the opportunities that exist to support social care and overcome some of its challenges;
- Treating social care change as a complex social-political problem that artificial intelligence is only a partial solution to;
- Encouraging and supporting new forms of social enterprises and/or business models to deliver artificial intelligence technologies to social care and healthcare.

Moreover, black-box artificial intelligence can deny people access to information generated from their own datasets – a risk to the ethical ownership of people’s data.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable? When should it not be permissible?

SCAMPI argues that, for the healthcare domain, black boxing is not acceptable. The details about how outcomes are computed, and thus can be explained, matters for most applications of artificial intelligence in this area.

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

The technology enabled care sector is largely unregulated and is only partially subject to industry standards. Therefore, there is a place for regulation in the uses of artificial intelligence in the delivery of health and social care. In England the Care Quality Commission has already considered a number of artificial intelligence-based services, mostly in primary care. Care regulators in each of the home nations need to be equipped with the necessary skills and capabilities to provide the public assurance.

Regulation needs to be independent of technology change and focused on how risk is managed, safety assured and how the outcomes of people using services are fulfilled. However, where artificial intelligence is developed to support and cooperate with people who require care and support and with health and care staff regulators need to be able to understand the role the artificial intelligence is playing in care processes and outcomes.

Currently, there is some ambiguity about where the contribution of industry standards end and statutory regulation start. Regulators need to future proof the way they regulate. The changing landscape needs to be mapped against the scope Parliament has determined for each relevant regulator. This would provide

assurance that there are no unintended gaps in what is regulated and that the responsibilities of different regulators are clear.

References

1. DeepMind Health, 2017, 'Helping clinicians get patients from test to treatment, faster', <https://deepmind.com/applied/deepmind-health/>
2. Kubota T., 2017, 'Deep learning algorithm does as well as dermatologists in identifying skin cancer', <http://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/>
3. Kirby J, 2017, 'Artificial intelligence better than scientists at choosing successful IVF embryos', <http://www.independent.co.uk/news/health/ai-ivf-embryos-better-scientists-selection-a7823736.html>
4. Kirtley A. & Maiden N.A.M., 2016, 'Creative Collaborations: the Care'N'Share app', *The Journal of Dementia Care* 24(2), 18-20.
5. Cellscope, 2017, 'Smarter Family Care' <https://www.celloscope.com/>
6. AliveCor, 2017, 'Meet Kardia Mobile: Your Personal EKG', <https://www.alivecor.com/>
7. Hart R., 2017, 'If you're not a white male, artificial intelligence's use in healthcare could be dangerous', <https://qz.com/1023448/if-youre-not-a-white-male-artificial-intelligences-use-in-healthcare-could-be-dangerous/>
8. Babylon Health, 2017, <https://www.babylonhealth.com/>
9. Cerri, Karin H., Knapp, Martin and Fernández, José-Luis, 2014, 'Decision making by NICE: examining the influences of evidence, process and context', *Health Economics, Policy and Law*, 9 (2). pp. 119-141. ISSN 1744-1331.
10. TIHM for Dementia, <http://www.sabp.nhs.uk/tihm>.
11. Burstow P., 2016, 'Social Care is running on Empty but technology can make a difference', <https://www.theguardian.com/social-care-network/2016/sep/13/social-care-running-on-empty-technology-paul-burstow>
12. IBM Watson Health, <https://www.ibm.com/watson/health/>

4 September 2017

Dr Valentina Rita Scotti, Dr Aysegul Bugra, Matthew Channon, Dr Ozlem Gurses and Dr Antonios Kouroutakis – Written evidence (AIC0051)

Dr Valentina Rita Scotti, Dr Aysegul Bugra, Matthew Channon, Dr Ozlem Gurses and Dr Antonios Kouroutakis – Written evidence (AIC0051)

[Submission to be found under Dr Aysegul Bugra](#)

Dr Huma Shah and Professor Kevin Warwick – Written evidence (AIC0066)

Introduction

1. Kevin Warwick and Huma Shah have been researching in artificial intelligence (AI) for over thirty five years. Warwick has written a lay book on the topic: *Artificial Intelligence: the Basics* (2011). Warwick and Shah have designed, organised and conducted original AI experiments involving members of the public as participants, including Lord John Sharkey acting as a Turing test Judge in a 2014 experiment. They have investigated the nature and effects of human-machine conversations realised from practical Turing test experiments and reported the results from these AI studies widely, including with transcripts of human-machine and human-human control tests, in academic journal papers, book chapters, and presented findings at international science conferences (Warwick & Shah, 2016ab; Warwick & Shah, 2014abc; Warwick & Shah, 2013; Warwick, et al., 2013; Shah, 2016; Shah, 2014; Shah, 2013; Shah, 2011; Shah & Warwick, 2016; Shah & Warwick, 2015; Shah & Warwick, 2010abc), and in a recent general readership book published by Cambridge University Press (2016c). The views expressed here are the authors founded in their AI scholarship.
2. In the sense that Alan Turing envisioned, in his scholarship on a machine's intellectual capacity (Turing, 1950), before and after his codebreaking at Bletchley Park, before his death in 1954, Artificial Intelligence (AI) does not yet exist. Co-opted by business and the healthcare industry, AI is considered a marketing term (Luminary Labs, 2017). Here we remind that Turing purposefully avoided defining intelligence, because as we know, defining words require other words which require explanations themselves. What Turing did was posit a methodology to explore the intellectual capacity of a machine through a comparison of its ability to answer any questions, as we have come to expect from our hands-free, voice activated digital assistants like Amazon's *Alexa* on its Echo and Echo Dot speaker, or Google's *Home*, two leaders in the home listening devices ready to answer home occupiers any questions with increasing accuracy, such as everyday requests to be prepared for the weather, '*Will I need an umbrella today?*' to general knowledge, '*how far away is Uruguay from the UK?*', and our interest in learning '*What causes hurricanes?*'. Amazon's *Alexa* assistant searches and provides answers to questions using Microsoft's **Bing** browser, while Google's *Home* uses the **Chrome** browser to find answers to users' queries.

Definition for Artificial Intelligence

3. There is nothing artificial about a machine successfully completing a task as well as, if not better than a human completing that task. The phrase was created by John McCarthy in 1955 in his proposal for a research conference which was held at Dartmouth College, US in 1956 (Stanford News, 2011). We prefer Turing's term – machine intelligence. In Turing's 1948 essay *Intelligent Machinery* (Turing, 1948), which philosopher Jack Copeland characterises as "the first manifesto of artificial intelligence" (Copeland, 2005: p.401), in which he proposed to investigate "the question as to whether it is possible for machinery to show intelligent behaviour" (in Copeland, 2005: p. 410), Turing presciently posited these five "suitable branches of thought for the machine to exercise its powers in":
 - i) Various games e.g. chess, noughts and crosses, bridge, poker
 - ii) The Learning of languages
 - iii) Translation of languages
 - iv) Cryptography
 - v) Mathematics.
4. Of the learning of languages, Turing (1948) stated "the learning of languages would be the most impressive, since it is the most human of these activities" (in Copeland, 2005: p. 421). Thus, here we bow to Turing's sagacity and omniscience to **define an artificial intellect as one being able to learn and talk with humans in human languages**. The fact that this innovation is not easy is borne out in evidence through social, customer service robots now sharing our public places, such as the *Spencer robot* scanning boarding passes at Amsterdam's Schiphol airport (University of Twente, 2016), or the security robots, and *Pepper* information robots in Westfield shopping mall in San Jose (Evangelista, 2016), and the robot airport guides providing travellers with information in Mineta San Jose airport (Petrova, 2016), both venues in Silicon Valley, California. These types of robots are providing some information and amusement for humans who interact with them, but they cannot yet hold a conversation with us humans, as do talking robots in umpteen cinematic movies through film history, including the *Tars* robot in *Interstellar* (Nolan, 2014), or the menacing *HAL* computer programme from *2001: A Space Odyssey* (Kubrick, 1968).

Current state of artificial intelligence

5. Apart from successes by machines bettering humans in human vs. machine games (chess in 1997; GO in 2016 and Poker in 2017), we focus on the great strides that have been made in natural language processing and understanding, since Turing, and Joseph Weizenbaum *Eliza*

programme (1966), which was the first enabling humans to chat with a machine (Shah, et al., 2016). Speech recognition is now instantaneous – dictation software such as *Dragon Naturally Speaking* can now recognise and allow dictation of whole utterances without pause by the speaker, whereas in the past, over 20 years ago dictation software like *Dragon Dictate* enforced a pause between words during dictation allowing the software to match the sound captured from the speaker with a store of words, this followed training the speaker had to compete in order for that system to record the speaker's idiolect (individual speech pattern).

The sectors most, and least likely, to benefit from artificial intelligence

6. Healthcare, medical surgery, transportation, smart cities, farming, military, disaster prevention and management, finance, insurance, retail, and education are among the sectors that will benefit from applying intelligent machines with advanced machine learning algorithms, boosted through the 4th Industrial Revolution embedding the Internet of Things (IoT). We believe machines/ robots do need to take over dull, dangerous jobs like coal-mining still done by children and women in some parts of the world – this is the least likely because humans in those places are currently cheaper to employ.

Other countries

7. US- China – Russia – India – Japan, across Europe, among other nation economies, are all investing heavily in artificial intelligence.

Public perception and the impact of artificial intelligence on society

8. With so much of artificial intelligence newsworthy and in news item in the print media, across television and in the plethora of digital media platforms the term is now mainstream. The general public are acquainted with AI technologies, such as Apple's *Siri*, Microsoft's *Cortana*, Samsung smart phone's *Bixby* assistant, as well as using Amazon's *Alexa* digital assistant to control home lighting and heating. The public are most probably also aware of concerns about AI voiced by notable scientists such as Stephen Hawking who has stated that "AI will either be the best or worst thing for humanity" (Guardian, 2016). Indeed Cambridge University's *centre to study existential risk* (CSER, 2016), cites AI as one of the threats to humanity. We believe the hype and the perceived threats are necessary for discussion. As with any technology, they are used to add value to processes, making businesses and industries more efficient and helping to grow local, regional and national economies, but there will always be those who wish to apply new technologies in a nefarious way.

The role of government in ensuring ethical dimensions in artificial intelligence

9. We conclude and recommend that Education be at the forefront of any government policy in relation to ensuring ethical dimensions in artificial intelligence. As STEM Ambassadors with numerous volunteering activities visiting schools, colleges and inviting children with their papers to participate in our AI experiments, we strongly believe that AI as a subject should be embedded into the school curriculum from primary age. There are a number of reasons why we advocate this. Firstly, we need to 'democratise AI' – it needs more diversity. Co-founder of Microsoft Bill Gates has noted that "AI has a sea of dudes" problem noting the mainly male attendees at a conference (Bloomberg, 2016). There are not enough women in AI. Microsoft's major study investigating female interest in STEM, of which AI is now a part, explored the views of over 11,000 women aged 11-30 across Europe found that "most girls become interested in STEM at the age of eleven-and-a-half but this starts to wane by the age of 15" (Trotman, 2017). We do not know whether it is because females undergoing puberty around that age leads to them losing their passion in STEM, what we do advocate is initiatives to get mothers involved in supporting their daughters at this crucial time. Anecdotal evidence shows some females are dissuaded from STEM subjects, for example being asked by their mother why they are studying physics, a *boys' subject*.
10. We also advocate innovative informal science education that can capture hidden talent among those who are from socio-economically disadvantaged groups and might not have the resources to visit science museums or not encouraged to watch science programmes, such as the annual BBC 2 Star Gazing Live on space science. We believe adapting Finland's model of *bring activity to the learner*, with the arts and the humanities powering innovative ways to engage more in AI and STEM, is one solution to resolving the problem of the lack of women in AI and STEM, and "poor white men, in places such as north-east England" who are "the least likely to go to university than anywhere else in the UK" (Coughlan, 2017). **Finland's Bioaika** (or **BioTime**) is an interactive exhibition about the forest sector realised through an itinerant truck visiting schools, cities and villages throughout Finland. The project is produced by the science centre of the Finnish forests administration Metsähallitus (Metsa, 2017), Pilke (2017), and the science centre Tietomaa (2017).
11. Adapting Finland's *bring activity to the learner*, we believe mobile science exhibitions touring places where science engagement is low, taking exhibitions into shopping malls where most people congregate with

innovative activities involving, say *Lego* could help to boost interest in AI and take up in formal education by more citizens.

12. For this to happen and be a success, a wide range of AI stakeholders, who include STEM education and STEM industry stakeholders, must come together to ensure hidden talent among those who might not be engaged in science but who may find AI fascinating through the most visual form of AI – robots, is given access to opportunities to seek more about how to get involved in AI. Engineering and other science and technology professionals' skills gap cannot continue without a detrimental effect on the economy and impacting society negatively. Acute shortage is expected to be felt in the cyber security field; not enough trained cyber security professionals are progressing through formal education and who are essential to tackle future envisioned problems such as preventing the hacking of driverless delivery vehicles. Both authors are ready to get the AI boom started, boosting the UK to lead in new and raw talent collaborating to design the next generation of AI technologies.

References

- Bloomberg. 2016. Artificial Intelligence has a 'sea of dudes' problem. *Bloomberg Business Technology*. Accessed from here: <http://www.seattletimes.com/business/technology/artificial-intelligence-has-a-sea-of-dudes-problem/>
- Coughlan. S. 2017. 10 Charts that show the effect of tuition fees. *BBC Education*. Accessed from here: <http://www.bbc.co.uk/news/education-40511184>
- Cellan-Jones, R. (2017). Computing in schools – alarm bells over England's classes. *BBC News – Technology*. Accessible from: <http://www.bbc.co.uk/news/technology-40322796>
- Copeland, B.J. 2005. *The Essential Turing: the ideas that gave birth to the computer age*. Oxford University Press: Oxford, UK
- CSER. 2016. *Centre for the Study of Existential Risk*. Accessed from here: <http://cser.org/>
- Evangelista, B. 2016. Robots greet Westfield mall shoppers in San Francisco & San Jose. *SFGate*. Accessed from here: <http://www.sfgate.com/business/article/Robots-greet-Westfield-mall-shoppers-in-San-10631291.php>
- Luminary Labs. 2017. Slide 20 of Hype vs Reality: The AI Explainer. Available from: <http://www.slideshare.net/LuminaryLabs/hype-vs-reality-the-ai-explainer/20-20Artificial intelligence AI Marketing term>
- Guardian, 2016. Stephen Hawking on the threat from AI. Accessed from: <https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge>
- Kubrick, S. 1968. *2001: A Space Odyssey*. IMDB, accessed from here: http://www.imdb.com/title/tt0062622/?ref_=nv_sr_1
- Metsa. 2017. *Metsähallitus*. Accessed from: <http://www.metsa.fi/web/en>

- Nolan, C. 2014. *Interstellar*. IMDB, accessed from here: http://www.imdb.com/title/tt0816692/?ref=nm_sr_1
- Petrova. M. 2016. Three customer serviced robots land in San Jose airport. *PC World* <http://www.pcworld.com/article/3141458/technology-business/three-customer-service-robots-land-in-san-jose-airport.html>
- Pilke, 2017. *Pilke Science Centre*. Accessed from: <https://www.tiedekeskus-pilke.fi/en/>
- Shah, H., Warwick, K., Vallverdu, J. and Wu, D. (2016). Can Machines Talk? Comparison of Eliza with Modern Dialogue Systems. *Computers in Human Behavior*. Vol 58, 278-295
- Shah, H. (2016). Keynote: Why the Turing test is relevant today. *Skolkovo AI conference*, Moscow, Russia: 14 Nov 2016
<https://sk.ru/foundation/events/november2016/ai/>
- Shah, H. (2014). The Emotions of Alan Turing: The boy who explained Einstein's Theory of Relativity aged 15 ½ for his mother. *International Journal of Synthetic Emotions*, Vol. 5(1), pp. 23-30
- Shah, H. (2013). Conversation, Deception and Intelligence: Turing's question-answer Game. In S.B. Cooper, & J. van Leeuwen (Eds) *Alan Turing: His Work and Impact*, pp. 614-620. Elsevier.
- Shah, H. (2011). Turing's Misunderstood Imitation Game and IBM's Watson Success. *Towards a Comprehensive Intelligence Test – Reconsidering the Turing Test for the 21st Century*. AISB Convention, 5 April, York University, UK
- Shah, H. and Warwick, K. (2010). From the Buzzing in Turing's Head to Machine Intelligence Contests. *Towards a Comprehensive Intelligence Test (TCIT)* symposium part of AISB Convention, De Montfort University, UK, March 29-30
- Shah, H. and Warwick, K. (2010). Hidden Interlocutor Misidentification in Practical Turing Tests. *Minds and Machines*, Vol. 20 (3), pp. 441-454
- Shah, H. and Warwick, K. (2010). Testing Turing's five-minutes, parallel-paired imitation game. *Kybernetes Turing test Special Issue*, Vol. 39 (3), pp. 449-465
- Shah, H. 2011. *Deception-detection and machine intelligence in practical Turing tests*. PhD thesis. Reading University UK.
- Shah, H., Warwick, K., Bland, I.M. and Chapman, C.D. 2014. Fundamental Artificial Intelligence: Machine Performance in Practical Turing tests. *Proceedings of 6th International conference on agents and Artificial Intelligence (ICAART 2014)*, 6-8 March, Angers, France
- Shah, H. and Warwick, K. (2016). Still want to know who is the human response? *Communications of the ACM*. Vol 59(9), p9
- Shah, H. and Warwick, K. (2015). Human or Machine? *Communications of the ACM*. Vol 58(4), p.8
- Stanford News. 2011. *Stanford's John McCarthy, seminal figure in artificial intelligence, dies at 84*. Accessed from here: <http://news.stanford.edu/news/2011/october/john-mccarthy-obit-102511.html>

- Tietomaa. 2017. *Teitomaa: Finland's first Knowledge Centre*. Accessed from: <http://www.tietomaa.fi/>
- Trotman, A. 2017. Why don't European girls like science or technology? *Microsoft Research*. Accessed from here: <https://news.microsoft.com/europe/features/dont-european-girls-like-science-technology/#sm.0000lmt9a4sx4e1xzzc1tfw5iphf4#fbFKuMIzRmZluYjB.97>
- Turing, A.M. 1950. Computing Machinery and Intelligence. *Mind*, Vol. 59(236), 433-460
- Turing, A.M. 1948. Intelligent Machinery. In (Ed) B.J. Copeland, *The Essential Turing: the ideas that gave birth to the computer age*, 2005. Oxford University Press: Oxford, UK
- University of Twente. 2016. *Robot Spencer accompanies first passengers at Schiphol airport*. Accessed from here: <https://www.utwente.nl/en/news/!/2016/4/497772/robot-spencer-accompanies-first-passengers-at-schiphol-airport>
- Warwick, K., and Shah, H. 2016a. *Turing's Imitation Game: Conversations with the Unknown*. Cambridge University Press: Cambridge, UK
- Warwick, K. and Shah, H. (2016b). Passing The Turing test does not mean the End of Humanity. *Cognitive Computation*. Vol. 8, pp 409-419
- Warwick, K. and Shah, H. (2016c). Taking the Fifth Amendment in Turing's Imitation Game. *Journal of Experimental and Theoretical AI*. DOI: 10.1080/0952813X.2015.1132273
- Warwick, K. and Shah, H. (2014c). Human Misidentification in Turing Tests. *Journal of Experimental and Theoretical AI*. Vol. 27(2) pp 123-135
- Warwick, K. and Shah, H. (2014b). Assumption of Knowledge and the Chinese Room in the Turing test. *AI Communications*. Vol. 27(3), 275-283
- Warwick, K. and Shah, H. (2014a). Effects of Lying in Practical Turing Tests. *AI & Society*, Vol. 31 (1) pp 5-15
- Warwick, K. and Shah, H. (2013). Good Machine Performance in Practical Turing tests. *IEEE Transactions on Computational Intelligence and AI in Games*. DOI: [10.1109/TCIAIG.2013.2283538](https://doi.org/10.1109/TCIAIG.2013.2283538)
- Warwick, K., Shah, H. and Moor, J. (2013). Some Implications of a Sample of Turing Tests. *Minds and Machines*, Vol. 23 (2) pp 163-177
- Warwick, K. 2011. *Artificial Intelligence: the Basics*. Routledge: Oxon, UK
- Weizenbaum, J. 1966. ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9 (1), 36-45

4 September 2017

Professor Noel Sharkey – Supplementary written evidence (AIC0248)

UK and Definitions of Autonomous Weapons Systems

A short report for the House of Lords select committee on Artificial Intelligence
Noel Sharkey

The UK Ministry of Defence definitions of Autonomous Weapons Systems (AWS), also known as Lethal Autonomous Weapons Systems (LAWS), and their interpretation of Automated Weapons Systems are out of step and at odds with how European and US allies and others are describing them at United Nations meetings such as the CCW. The definitions are also at odds with the engineering community.

Part 1 of this brief report outlines the definitional differences between the UK and its allies on Autonomous Weapons Systems. Part 2 focuses on how MoD documents assign the dividing lines between Automated and Autonomous Weapons Systems differently than others. This creates a definitional conflation that clouds political judgments and impacts negatively on the UK's ability to develop coherent policies on autonomy in weapons that are consistent with and relevant to the international community of nations. Part 3 considers the need for a definition of Autonomous Weapons Systems that includes the type of human control required for compliance with International Law. There is an opportunity here for the UK to 'get ahead of the game' and show international leadership on the issue.

1. Autonomous Weapons Systems¹²⁰⁴

According to the UK Ministry of Defence, two of the requirements for AWS status are that they must be:

(i)"self-aware and their response to inputs indistinguishable from, or even superior to, that of a manned aircraft. As such, they must be capable of achieving the same level of situational understanding as a human."¹²⁰⁵

(ii)"capable of understanding higher level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives."¹²⁰⁶

¹²⁰⁴ Thanks to Daan Kayser from PAX Netherlands for assistance in compiling the European Definitions. See also PAX report Keeping Control, October 2017 www.paxforpeace.nl

¹²⁰⁵ UK Ministry of Defence, Development, Concepts and Doctrine Centre (2011) The UK Approach to Unmanned Aircraft Systems, Joint Doctrine Note, 30 March 2011

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/33711/20110505J

¹²⁰⁶ Ministry of Defence, 'Joint Doctrine Publication 0-30.2: unmanned aircraft systems' September 2017, <https://www.gov.uk/government/publications/unmanned-aircraft-systems-jd-0-302>.

These machines are unlikely to exist in the near future if ever. As the MoD correctly point out, “machines with the ability to understand higher-level intent, being capable of deciding a course of action without depending on human oversight and control currently do not exist and are unlikely in the near future.”

Such ‘science fiction’ requirements can misdirect the UK into inferences such as, ‘since they are unlikely to exist in the near future, we do not need to consider their impact on the nature of armed conflict or consider prohibiting or regulating them.’ Others define AWS in a realistic way that is consistent with new developments in weaponry in the hi-tech nations including the UK.

In the field of robotics the terms autonomy and autonomous robot are used with specific meanings that are only vaguely related to the political and philosophical definitions of autonomy. They were first used to indicate that the robot had an onboard computer (when computers got small enough). An autonomous robot is a mobile robot that can perform tasks in an (usually) unstructured environment without human supervision or guidance. Sensors on the robot send information to a computer or controller that operates motors to perform the tasks. A good example is the Roomba vacuum cleaning robot. In contrast, a semi-autonomous robot can perform some of its tasks without human intervention. This differs from an automatic robot that carries out a set of preprogrammed and predefined actions in a fixed environment e.g. painting a car.

The key component of an autonomous weapons system is that it has autonomy in the critical functions of target selection and the application of violent force. In other words, a weapons systems that can select targets and apply force without human supervision at the time of attack. This is how AWS are discussed at the UN. Below are significant extracts from the definitions of European state actors, the US and the International Committee of the Red Cross that evidence this.

US “A weapon system that, once activated, can select and engage targets without further intervention by a human operator.”¹²⁰⁷

International Committee of the Red Cross (ICRC) “Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention.”¹²⁰⁸

France “[LAWS imply] a total absence of human supervision, meaning there is absolutely no link (communication or control) with the military chain of command

¹²⁰⁷ US Department of Defense (DoD), Autonomy in Weapon Systems, Directive 3000.09, 21 November 2012 and its amended version (still Directive 3000.09) 2017

¹²⁰⁸ Views of the International Committee of the Red Cross on autonomous weapon systems at the CCW Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS) 11-15 April 2016, Geneva

... targeting and firing a lethal effector (bullet, missile, bomb, etc.) without any kind of human intervention or validation.”¹²⁰⁹

Norway: “weapons that would search for, identify and attack targets, including human beings, using lethal force without any human operator intervening.”¹²¹⁰

Austria “[AWS are] weapons that in contrast to traditional inert arms, are capable of functioning with a lesser degree of human manipulation and control, or none at all.”¹²¹¹

Italy “[Lethal AWS are systems that make] autonomous decisions based on their own learning and rules, and that can adapt to changing environments independently of any pre-programming” and they could “select targets and decide when to use force, [and] would be entirely beyond human control,”¹²¹²

Switzerland “[AWS are] weapons systems that are capable of carrying out tasks governed by IHL in partial or full replacement of a human in the use of force, notably in the targeting cycle.”¹²¹³

The Netherlands “a weapon that, without human intervention, selects and attacks targets matching certain predefined characteristics, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention.”¹²¹⁴

The Holy See “An autonomous weapon system is a weapon system capable of identifying, selecting and triggering action on a target without human supervision.”¹²¹⁵

¹²⁰⁹ Working Paper of France, ‘Characterization of a LAWS’, CCW informal meeting of experts on LAWS, Geneva, April 2016,

¹²¹⁰ Statement of Norway, CCW informal meeting of experts on LAWS Geneva, 13 April 2016, http://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2016/meeting-expertslaws/statements/13April_Norway.pdf

¹²¹¹ Statement of Austria, CCW informal meeting of experts on LAWS, Geneva, 13 May 2014, <https://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/assets/media/22D8D3D0ACB39CA8C1257CD70044524B/file/Austria%20BMX%20BLAWS.pdf>.

¹²¹² Statement of Italy, CCW informal meeting of experts on LAWS, Geneva, 12 April 2016

¹²¹³ Informal Working Paper submitted by Switzerland, CCW informal meeting of experts on LAWS, 30 March 2016, <http://www.reachingcriticalwill.org/images/documents/Disarmament-fora/ccw/2016/meeting-expertslaws/documents/Switzerland-compliance.pdf>

¹²¹⁴ AIV/CAVV, ‘Autonomous weapon systems; the need for meaningful control’, October 2015, a synopsis of the report can be found here: <http://aiv-advies.nl/8gr#advice-summary> and the full report here <http://aivadvies.nl/download/606cb3b1-a800-4f8a-936f-af61ac991dd0.pdf>.

¹²¹⁵ Working Paper submitted by the Holy See, ‘Elements Supporting the Prohibition of Lethal Autonomous Weapons Systems’, April 2016, <http://www.reachingcriticalwill.org/images/documents/Disarmamentfora/ccw/2016/meeting-experts-laws/documents/HolySee-prohibition-laws.pdf>.

2 Automated v Autonomous Weapons Systems

An important issue in UN discussions about Autonomous Weapons Systems is that there are number of weapons currently being used for high speed defences such as shooting down missiles, mortar shells and swarm attacks on ships. Examples of weapons include C-RAM, Phalanx, NBS Mantis and Iron Dome. These systems complete their detection, evaluation and response process within a matter of seconds and thus render it extremely difficult for human operators to exercise meaningful supervisory control once they have been activated other than deciding when to switch them off.

There is understandable concern that new regulations or prohibitions of autonomous weapons, may impact on use of such defensive weapons. Thus it is felt that there is a definitional need to separate these from Autonomy in weapons systems.

Some suggest calling the defensive weapons systems *automatic* or *automated* rather than *autonomous*. For example, the International Committee of the Red Cross proposes that, "An automated weapon or weapons system is one that is able to function in a self-contained and independent manner although its employment may initially be deployed or directed by a human operator."¹²¹⁶ The US Department of Defence suggests that, "...the automatic system is not able to initially define the path according to some given goal or to choose the goal that is dictating its path."¹²¹⁷ There may also be other ways to make this separation between defensive and autonomous weapons systems (See Appendix 1 on SARMO systems).

UK definitions conflate autonomous and automated weapons systems: The UK definition is similar to those of others: "... an automated or automatic system is one that, in response to inputs from one or more sensors, is programmed to logically follow a predefined set of rules in order to provide an outcome. Knowing the set of rules under which it is operating means that its output is predictable."¹²¹⁸ However, MoD pushes autonomous weapons systems into the category of automated weapons systems. Sometimes they refer to these as advanced automation or highly automated. This conflates autonomous and automated weapons systems in contrast to the definitions of other nation states.

This definitional conflation leads to UK thinking about Lethal Autonomous Weapons Systems (LAWS) that is out of step with their allies "the UK believes that LAWS do not, and may never, exist." Yet according to the definitions of others the US, China, Israel and Russia are the front-runners in developing and testing prototype

¹²¹⁶ ICRC (2011) International Humanitarian Law and the challenges of contemporary armed conflicts, p 39

¹²¹⁷ US Department of Defense (2013) Unmanned Systems Integrated Roadmap, FY2013-2038, p 66

¹²¹⁸ It is important to note here that when a mobile device is being controlled by information detected by sensors, its exact behavior cannot be predicted in an open ended or unstructured environment.

autonomous tanks, fighter jets, submarines, ships and swarm technology. Swarm technology is about the creation of force multiplication with large numbers of attack vehicles operating autonomously together.

The reason why the UK takes a stance on saying that LAWS may never exist is merely definitional. By setting an unrealistic requirement for its definition of LAWS it places them into the category of automated weapons. This hides, either inadvertently or deliberately, the UK's views and plans for autonomous weapons. This shows up most in the MoD document *Future Operating Environment 2035* (FOE35)¹²¹⁹ where the term automated weapons systems is used to refer to what others call Autonomous Weapons Systems or Lethal Autonomous Weapons Systems.

Thus whilst the UK continue to say that they will never develop autonomous weapons systems and thus do not see the need to support new regulations, a moratorium or a prohibition on them, they use veiled comments in FOE35 to say that, "our immediate priorities should be: ... investment in emerging technologies, especially automated systems." (p30) and "Defence will need to make exploiting emerging technology and capability in automated systems a priority, as well as countering our opponents' systems." (p32)

It is clear that the UK, as a morally upstanding nation, has concerns about the ethical and legal use of what they call advanced automated systems (read Autonomous Weapons Systems): "Our legal and societal norms will continue to apply restraint to the conduct of military operations, particularly violent conflict, out to 2035. This will be particularly true where this applies to new technologies such as automated systems and novel weapons.

And they also show engagement with some of the arguments against and problems with LAWS at the UN. But they do not articulate these concerns because they all fall under the UK definition of automated weapons systems. Here are three quotes from FOE35 to demonstrate that concerns about the impact of LAWS on global security are being hidden under the term 'automated'.

(i) **Use of AWS by rogue nations and non-state actors:** "Our potential adversaries may not be so constrained, and may operate without restraint." (p44) and "in the virtual environment, swarm attacks could be planned through crowd-sourcing before being executed through multiple access points in multiple countries, making deterrence and defence against them almost impossible. These could be orchestrated by terrorists," (p41)

(ii) **Proliferation of AWS:** "Automated systems, including those that are armed, will proliferate over the next 20 years. Advances in technology will almost certainly enable swarm attacks, allowing numerous devices to act in concert. This may serve

¹²¹⁹ UK Ministry of Defence, Strategic Trends Programme: Future Operating Environments 2035, 14 December 2015
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/646821/20151203-FOE_35_final_v29_web.pdf

to counter the advantage of high-end systems.” (p27) and “As they become cheaper and easier to produce, technologically advanced systems are likely to proliferate, with developing states and non-state actors having growing access to capable systems.” (p16)

(iii) **Lowering the threshold for resorting to violent conflict:** “change the threshold for the use of force. Fewer casualties may lower political risk and any public reticence for a military response.”

The UK has rendered itself unable to make these arguments about AWS at the UN because it has dismissed them out of hand by giving AWS a science fiction definition and hiding what everyone else calls AWS or LAWS under the cloak of automated weapons systems.

3. Human Supervisory Control of weapons systems

One necessary requirement for defining a weapons system as autonomous is common to all: autonomous weapons systems are weapons that operate without human control. The hub of the debate on autonomous weapons systems concerns what is meant by human control. While all Nation States say that their weapons will be under human control, they do not specify what this mean.

The Parliamentary under secretary of state, Lord Astor of Hever, stated that: “[T]he MoD currently has no intention of developing systems that operate without human intervention ... let us be absolutely clear that the operation of weapons systems will always be under human control”.¹²²⁰ What has not been made absolutely clear in the United Kingdom, however, is exactly what type of human control will be employed.

To say that there is a human in the control loop does not clarify the degree of human involvement. It could simply mean a human programming a weapons system for a mission or pressing a button to activate it, or it could (hopefully) mean exercising full human judgment about the legitimacy of a target before initiating an attack. The UK NGO Article 36 coined the term *meaningful human control*¹²²¹ to facilitate discussions about the type of human control for every attack that is acceptable under international law.

The UK has a wonderful opportunity here to show international leadership by laying out in detail what human control of weapons systems means. Clues to the type of process required for this analysis and the problems are provided in Appendix 2. This is not intended to be treated as dogma. It has been derived from the large scientific literature on human supervisory control of machinery.

¹²²⁰ 26 March 2013. Cf. <http://bit.ly/1IZMQyW> _14.

¹²²¹ Article 36, Autonomous weapons, meaningful human control and the CCW, May 21, 2014 <http://www.article36.org/weapons-review/autonomous-weapons-meaningful-human-control-and-the-ccw/>

APPENDIX 1: A draft definition of defensive weapons

There are currently weapons systems in use that operate automatically once activated. Such SARMO (Sense and React to Military Objects¹²²²) weapon systems intercept high-speed inanimate objects such as incoming missiles, artillery shells and mortar grenades automatically. Examples include C-RAM, Phalanx, NBS Mantis and Iron Dome. These systems complete their detection, evaluation and response process within a matter of seconds and thus render it extremely difficult for human operators to exercise meaningful supervisory control once they have been activated other than deciding when to switch them off.

There are a number of common features for SARMO weapons¹²²³ that are necessary although not sufficient to keep them within legal bounds:

- fully pre-programmed to automatically perform a small set of defined actions repeatedly and independently of external influence or control
- used in highly structured and predictable environments that are relatively uncluttered with very low risk of civilian harm
- fixed base – although these are used on manned naval vessels, they are fixed base in the same sense as a robot arm on a ship would be.
- switched on after detection of a specific threat
- unable to dynamically initiate a new targeting goal or change mode of operation once activated
- have constant vigilant human evaluation and monitoring for rapid shutdown in cases of targeting errors, change of situation or change in status of targets
- the output and behaviour of the system is predictable
- only used defensively against direct attacks by military objects

The US Department of Defense calls these human supervised autonomous weapons: "Human-supervised autonomous weapon systems may be used to select

¹²²² The term SARMO weapons first appeared in, Sharkey N. Towards a principle for the human supervisory control of robot weapons', *Politica and Società*, 2 (2014), 305–24.

¹²²³ Fire and forget weapons such as radiation detection loitering munitions and heat seeking missiles are not included here and require a separate discussion.

and engage targets, with the exception of selecting humans as targets, for local defense to intercept attempted time-critical or saturation attacks for:

- (a) Static defense of manned installations.
- (b) Onboard defense of manned platforms.¹²²⁴

It is the human decision of when to use the weapon that is key to the legality of SARMO weapons systems. It is essential for making such decisions that precautionary measures have been taken about the target's significance - its necessity and appropriateness, and likely incidental and possible accidental effects of the attack¹²²⁵. It is also essential that vigilance is maintained during operation of the weapons systems and that there is a means for rapidly deactivating the weapons if it becomes apparent that the objective is not a military one or that the attack may be expected to cause incidental loss of civilian life.¹²²⁶

APPENDIX 2 Ideas for the analysis of human control of weapons

In order to ensure the legality of human control of weapons it is necessary to ensure that any interface between operators and weapons is designed with an understanding of human psychological processes. This is required to guarantee that precautionary measures are taken about the significance of a target, its necessity and appropriateness, and likely incidental and possible accidental effects of the attack.

To see how this works, we can look at fundamental types of control as shown in Table 1.¹²²⁷

Table 1: A classification for levels of human supervisory control of weapons
human deliberates about a target before initiating any attack
program provides a list of targets and human chooses which to attack
program selects target and human must approve before attack
program selects target and human has restricted time to veto
program selects target and initiates attack without human involvement

¹²²⁴ US Department of Defense (2012) op cit. p7

¹²²⁵ As specified Article 57 of additional protocol 1 to the Geneva Convention 1977
<http://bitly.com/1hJF4GC> Last accessed March 5 2014

¹²²⁶ See article 57 2(iii)(b) *op cit* for a full account.

¹²²⁷ For a more in-depth understanding of these analyses see Sharkey, N.E. (2016) *Staying in the Loop: human supervisory control of weapons*, in Bhuta Nehal and Hin Yan Lui (eds), *Autonomous Weapons Systems and the Law*, Cambridge University Press.

Level 1 control is the ideal. A human commander (or operator) must have full contextual and situational awareness of the target area at the time of a specific attack and be able to perceive and react to any change or unanticipated situations that may have arisen since planning the attack. There must be active cognitive participation in the attack and sufficient time for deliberation on the nature of the target, its significance in terms of the necessity and appropriateness of the attack, and likely incidental and possible accidental effects. There must also be a means for the rapid suspension or abortion of the attack.

Level 2 control could be acceptable if shown to meet the requirement of deliberating on the potential targets. The human operator or commander should be in a position to assess whether an attack is necessary and appropriate, whether all (or indeed any) of the suggested alternatives are permissible objects of attack, and to select the target which may be expected to cause the least civilian harm. This requires deliberative reasoning. Without sufficient time or in a distracting environment the illegitimacy of a target could be overlooked.

A rank ordered list of targets is particularly problematic as there would be a tendency to accept the top ranked target unless sufficient time and attentional space is given for deliberative reasoning.

Level 3 is unacceptable. This type of control has been experimentally shown to create what is known as *automation bias* in which human operators come to trust computer generated solutions as correct and disregard or don't search for contradictory information. Cummings experimented with automation bias in a study on an interface designed for supervision and resource allocation of in-flight GPS guided Tomahawk missile.¹²²⁸ She found that when the computer recommendations were wrong, operators using Level 3 control had a significantly decreased accuracy.

Level 4 is unacceptable because it does not promote target identification and a short time to veto would reinforce automation bias and leave no room for doubt or deliberation. As the attack will take place *unless* a human intervenes, this undermines well-established presumptions under international humanitarian law that promote civilian protection.

The time pressure will result in operators neglecting ambiguity and suppressing doubt, inferring and inventing causes and intentions, being biased to believe and confirm, focusing on existing evidence and ignoring absent but needed evidence. An example of the errors caused by fast veto came in the 2004 war with Iraq when the U.S. Army's Patriot missile system engaged in fratricide, shooting down a British Tornado and an American F/A-18, killing three pilots.¹²²⁹

¹²²⁸ Cummings, ML (2006) Automation and Accountability in Decision Support System Interface Design, *Journal of Technology Studies*, vol. 32, 23-31

¹²²⁹ Cummings (2006) *ibid*

Level 5 control means control by computer alone and it is therefore refers to an autonomous weapons systems.

It should be clear from the above that research is urgently needed to ensure that human supervisory interfaces make provisions to get the best level of human control needed to comply with the laws of war in all circumstances.

9 February 2018

Simul Systems Ltd – Written evidence (AIC0016)

This paper addresses a number of items from the Call for Evidence: namely that on The Pace of Technical Change, Ethics, the Role of Government, and Overseas Legislative Sources.

Author: Andrew J Lewis MSc FLS, Director

1) The change in the technologies of production of AI.
A major feature of rule-based expert systems is that they are not programmed using procedural scripts, stepping from one stage to another in the way of a BASIC or COBOL program, but Expert Systems¹²³⁰ using logic languages such as PROLOG describe logical constraints and seek goals. The precise manner in which the program performs this goal-seeking behaviour is not always very visible, can be huge and very complex, particularly if using heterogeneous data as a basis. It is possible to output to a log file the status of the system at any given stage in the process, but these log files are likely to be huge and incomprehensible except to experts. The maintenance of huge log files will be a large cost to the user organization.

With machine learning systems the process is data-driven, not program driven. Thus it may well be almost impossible to predict and therefore test for all the scenarios the system may face? In the case of really Big Data the log files for processing would be arcane to decipher even for an expert. This may also be true of Expert Systems?

2) Our experience and concerns
We are currently working on the development of an Artificial Moral Agent program, called Project ERGON, and have developed a first draft prototype called "Judgementis", a system which makes judgments on outcomes in Human Relations Disciplinary cases. So far the reasonableness of the decisions is encouraging.

With regard to product liability, is there scope for a cap or ceiling on such liability to ensure insurance for SMEs is not prohibitive to innovation and growth?

3) The Lingua Franca of explanations.
The legal requirement in the GDPR¹²³¹ legislation and other requests from the EU Parliament for systems to explain their decisions in cases where humans are affected and involved means that there needs to be a standard lingua franca of outputs agreed internationally to make sure that

¹²³⁰ Expert System – a branch of Artificial Intelligence which develops adviser systems

¹²³¹ EU General Data Protection Regulation

non-experts, courts and media can understand what went on. The other side of this is that every system would have to have as a specific component a huge translation module, absorbing processing time, just to record history? It would not be impossible to develop such a lingua franca, or translation table from computer codes to English, for example, but would require a joint design committee formed from members of the judiciary, insurance, developers including micro-SMEs¹²³², legislators and representatives of the public, “the person on the Clapham Omnibus”. It would also enable a discussion about the potential for the explanation to expose the developer’s Intellectual Property intrinsic in the programming to their competitors, to the detriment of orderly commerce?

I heartily recommend that such a committee be formed to agree the parameters, design and use of such a translation function for decision explanation, a ‘lingua franca’.

4) On Ethics

There is a fundamental philosophical paradigm which in our opinion needs to be considered when implementing, monitoring or controlling Artificial Intelligence programs. That is, how does the legal system view the rights, duties and expectations of the machine, and are the laws sufficiently knowledgeable of the domain to be fair and effective? For example, in our research we have identified hypothetical ethical dilemmas which require the survival or well being of the principal actor to be considered, such as in an emergency? In these cases the individual can either act in their own interest or altruistically. However a post-event judgment would weigh whether that individual person should consider their own safety, whether it was “reasonable” so to do?

What however if the actor were a machine? Would survival, altruism and even love be part of the calculus? Should we expect a machine to behave morally like a person, or a machine? With current states of the art, such considerations would be difficult and expensive to program. Furthermore they make the point that we view people and machines differently, their relative “cleverness” notwithstanding? Therefore laws which include ethics and morality for people may not always be wholly appropriate for a machine? But then neither may Product Liability laws which are machine-centric? We argue that Law-makers should reconsider Public and Product Liability laws with reference to the point above to enact amendments which cater for two sorts of actors, people and machines.

As a further complication, there are two sorts of ethics at play here: the ethics of program function, and of implementation. The first is either intrinsic or explicit, where the system is programmed to obey a set of rules which adhere to the programmer’s interpretation of ethics, and next

¹²³²Micro-SME - Very small and medium sized enterprises including individuals; “one-person bands”.

where the program itself implements ethical decision making? The second sort of ethics is society's response to the technological challenge posed by the onward rush of AI. Both of these ethical paradigms should be considered on a machine-centric or people-centric basis as above?

5) The role of Government

The role of the UK Government is complex from our perspective. It must satisfy democratic demands for control and transparency, as with GDPR, but as we are discovering with that legislation, there is a danger of over-reaction and hampering business opportunities and commercial wellbeing? It must protect and yet encourage, simultaneously. However the key to a thriving software development sector is knowing what is going on at this critical time?

To that end, this Committee might urgently consider establishing a conduit via the Parliamentary website, appropriate Government department, or the FSB¹²³³ or BCS¹²³⁴ to disseminate relevant information about new laws and amendments being considered by Parliament? At present such information is hard to come by and usually the result of chance. There should be a means for new tech SMEs to be a part of the design and implementation process, and for them to know what to do? This has been a lesson from GDPR¹²³⁵ - a surprise law which as yet provides no framework for developers operating in advance of the current markets to incorporate design changes to their software products to conform to legislation. We need agreed standards and specifications in good time?

6) Overseas Legislative sources

With regards to overseas legislation, we are already implementing EU Regulations, a process of which we are shortly likely to be no longer party to. However we are likely to remain major trading partners and should monitor EU legislation carefully. We must also monitor USA, Australian, New Zealand and Indian developments as these are instantiated in similar legal systems to our own, and Parliament should also be monitoring the Japanese and South Korean endeavours which may well surpass our own? However the key issue is how we cater for the announced massive Chinese effort? There may be complications for AI Ethics here as the Chinese Communist Party ethics may be somewhat different in places to Western Liberal thought, and there may be a need again for a joint Committee to develop a "**shared bedrock of values**"¹²³⁶ for developers to follow in software products?

¹²³³ The Federation of Small Businesses

¹²³⁴ British Computer Society

¹²³⁵ Ibid.

¹²³⁶ Attrib. to Professor Alan Fox, Oxford School of Industrial Relations from a North East London Polytechnic tutorial.

17 August 2017

Jonathan Sinclair – Written evidence (AIC0023)

Question 1: The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this?

1.1 The field of Artificial Intelligence is enjoying a renaissance which has picked up pace in the last 5 years. Achieving successes in domains once believed intractable for at least the next 20 years, now having been solved¹²³⁷, and despite some commentators believing that what we're seeing is simply a refinement of expert systems, many believe a revolution in this technological space is taking place.¹²³⁸

The results of these advancements are down to algorithmic and computational hardware improvements that have seen specialized light-weight computational units (Graphical Processing Units (GPU's), Floating Point Gate Arrays (FPGA's)) manage distributed computational problem dissemination, at scale.

To understand the current state-of-the-art a few of the key advancements should be mentioned, which bring into focus the areas where AI is enjoying continued success:

- 2010: Microsoft Kinect device released which demonstrated an advancement in computer vision being able to track a person's body movements accurately and in real-time¹²³⁹
- 2011: IBM's Watson computational system wins the game of Jeopardy against human champions¹²⁴⁰
- 2015: Google DeepMind's AlphaGo application defeats world's 2nd ranked human Go champion¹²⁴¹
- 2017: Carnegie Mellon's Libratus application defeats 4 top players at no-limit Texas hold 'em Poker¹²⁴²

Excluding the first instance, a point could be made that advances in AI are only enjoying success at game-playing problem solving tasks while a counter

¹²³⁷ CNN: Elon Musk backs call for global ban on killer robots (2017) available at: <http://money.cnn.com/2017/08/21/technology/elon-musk-killer-robot-un-ban/index.html>, accessed 22.08.2017

¹²³⁸ The Fourth Industrial Revolution: what it means, how to respond (2016,) available at: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>), accessed 22.08.2017

¹²³⁹ Wikipedia Kinect entry (2017), available at: <https://en.wikipedia.org/wiki/Kinect>, accessed 22.08.2017

¹²⁴⁰ The Guardian: IBM computer Watson wins Jeopardy clash, (2011), available at: <https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy>, accessed 22.08.2017

¹²⁴¹ Scientific American: How the Computer Beat the Go Master, (2016), available at: <https://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/>, accessed 22.08.2017

¹²⁴² IEEE Spectrum: AI Decisively Defeats Human Poker Players, (2017), available at: <http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/ai-learns-from-mistakes-to-defeat-human-poker-players>, accessed 22.08.2017

suggestion could be proposed that all decision-making processes and ergo, intelligent types of problem solving activities, spawn from this type of logical inductive and deductive thinking.

These are not, however, the only areas where AI is helping shape our current societal fabric and although the previous exemplars may serve to demonstrate certain archetypical forms, nearly every other sector of society is at some level being touched by advances in this area. To give a few examples:

- Medicine: AI diagnoses cancer better than human pathologists¹²⁴³
- Self-driving (AI assisted) vehicles now permitted to operate along with human counterparts on open roads¹²⁴⁴
- Legal assistance, automated legal support¹²⁴⁵
- Financial Trading¹²⁴⁶
- Musical composition¹²⁴⁷

Every one of these problem domains not only relies on the aforementioned hardware items being in place to allow for scale and parallelization, they also require an incremental step-change concerning how computational devices are used to solve human problems.

Traditionally, algorithmic logic has followed a statically-based operational flow whereby commands are issued the “*If..then..else*” grammatical form e.g. If X happens then do Y, otherwise do Z. This procedural mechanism for solving problems requires that the programmer of the algorithm understands the problem space s/he is trying to solve, can describe the exact steps required to obtain a solution and enable it via a computational language.

Rediscovered statistical methods, that find realization as dynamically defined computational programs, give computers the ability to react and learn to given inputs and outputs in an unpredictable manner, meaning that the human operator is removed from having to describe every action required to get to a solution and can instead simply define a solution and ‘train’ a dynamic system to attain the goal.

The algorithmic sets currently in favor (and attributed to providing the startling success of recent times) are defined as Neural Networks, whose sub-domain of Deep-learning algorithms are ushering in and maintaining the recent fervor

¹²⁴³ Google’s Deep Learning AI Diagnoses Cancer Better Than Human Pathologists, (2017), available at: <https://edgylabs.com/google-ai-cancer-diagnosis/>, accessed 22.08.2017

¹²⁴⁴ SwissInfo: Which countries are testing driverless cars? (2016), available at: https://www.swissinfo.ch/eng/sci-tech/future-of-transport_which-countries-are-testing-driverless-cars-/41999484, accessed 22.08.2017

¹²⁴⁵ New York Times: A.I. Is Doing Legal Work. But It Won’t Replace Lawyers, Yet. (2017), available at: <https://www.nytimes.com/2017/03/19/technology/lawyers-artificial-intelligence.html?mcubz=3>, accessed 22.08.2017

¹²⁴⁶ Wired Business: The rise of the artificially intelligent hedge fund, (2016), available at: <https://www.wired.com/2016/01/the-rise-of-the-artificially-intelligent-hedge-fund/>, accessed 22.08.2017

¹²⁴⁷ Nvidia: AI Podcast: The Next Hans Zimmer? How AI May Create Music for Video Games, Exercise Routines (2017), available at: <https://blogs.nvidia.com/blog/2017/08/09/ai-podcast-ai-music-gtc-pierre-barreau/>, accessed 23.08.2017

around this field, other forms of dynamic programming technologies exist however haven't resonated to such effect so far.

Some commentators may suggest many other factors are contributing to the state of AI e.g. globalization pressures to compete against low cost work forces, the drive towards automation, renewed scalability requirements, increased complexity of handling the huge data-sets now available, the digital revolution, the human need to catalogue information etc. however these are all secondary in nature.

Without the pressure to deliver more computational power and advances in neural network technology, the current state of AI would still be in a mode of stagnation.

1.2 How is it likely to develop over the next 5, 10 and 20 years?

Making predictions for the future of this technology is extremely difficult given the enthusiasm and hiatus cycles previously experienced by the field. The nearer term predictions are always easier to get right than further out, however the following can be said:

- In the next 5 years, we can expect to see an increasing interest and dependence on AI-powered technologies given that all the major technological players in this area are starting to build their own hardware specifically to address this e.g. Apple¹²⁴⁸, Google¹²⁴⁹, Microsoft¹²⁵⁰.

This will lead to advancements in the following areas:

- o Computer Vision and object identification becomes comparable to human (Nvidia¹²⁵¹, autonomous vehicles, laser optics etc.)
- o Computer speech and translation services replace human operators¹²⁵²
- o Autonomous vehicles reach Level 5 of SAE Internationals automation scale, meaning full autonomy¹²⁵³
- o Military units become self-defending with autonomous capabilities
- o Fully-automated racing events will mature and gain a greater following¹²⁵⁴
- o Automated delivery mechanisms will be carried out by drones

¹²⁴⁸ Phys.org: Apple's new mobile AI chip could create a new level of intelligence, (2017), available at: <https://phys.org/news/2017-05-apple-mobile-ai-chip-intelligence.html>, accessed 22.08.2017

¹²⁴⁹ Wired Business: Google rattles the tech world with a new AI chip for all (2017), available at: <https://www.wired.com/2017/05/google-rattles-tech-world-new-ai-chip/>, accessed 22.08.2017

¹²⁵⁰ Techcrunch: Microsoft's second-generation HoloLens will include a dedicated AI coprocessor, (2017), available at: <https://techcrunch.com/2017/07/24/microsofts-second-generation-hololens-will-included-a-dedicated-ai-coprocessor/>, accessed 22.08.2017

¹²⁵¹ Nvidia: AI car computing platform, (2017), available at: <http://www.nvidia.com/object/drive-px.html>, accessed 22.08.2017

¹²⁵² Business Insider: Microsoft's AI is getting crazily good at speech recognition, (2017), available at: <http://uk.businessinsider.com/microsofts-speech-recognition-5-1-error-rate-human-level-accuracy-2017-8?r=US&IR=T>, accessed 22.08.2017

¹²⁵³ SAE International: Automated Driving, (2014), available at: https://www.sae.org/misc/pdfs/automated_driving.pdf, accessed 22.08.2017

¹²⁵⁴ Wikipedia Roborace entry, (2017), available at: <https://en.wikipedia.org/wiki/Roborace>, accessed 23.08.2017

- Growth of human-absent high-street stores will start to enter the mainstream shopping markets
- Manufacturing and industrial factories will see a reduction of human personnel
- In the next 10 years, we can expect existing technologies to mature to the point where a greater trust is given over to AI entities and the technology starts to blend into our reality with the advent of things like:
 - Fully predictive prosthetics
 - Mixed-reality models (leveraging back-end driven AI analytical engines) become ubiquitous
 - AI-driven interfaces dominate computational interaction, learn and adapt to your preferences
 - Increased robotic presence within society. Robotic entities become commonplace
 - Personalized medicine tailored to patients
 - Automated medical diagnosis applications
 - Mobile technology integrated with human biological makeup
 - All levels of societal employment will be augmented through AI technologies
 - The rise of an AI resistant sub-culture, causing frictions concerning ownership and accountability within society
 - Generalised AI models develop
 - Smart/Adaptive homes
- In the next 20 years one of two scenarios are likely to play out:
 - AI technological development will continue to exponentially increase with no abating, meaning that the term AI will probably still hold its mantle, however it will be crude and condescending to refer to AI as artificial and we'll start talking about intelligent entities/agents (IE/IA's), whose rights have to be recognized and a symbiosis will occur between humanity and technology.
 - The other scenario has AI not delivering on the present-day predications and it falling once again into an AI Winter¹²⁵⁵, whereby trust and generalization can't be achieved which leads to a stagnation in the field, testifying to existing worries that deep learning technologies have severe limits¹²⁵⁶.

1.2 What factors, technical or societal, will accelerate or hinder this development?

Several factors will either accelerate or hinder this development.

¹²⁵⁵ Wikipedia AI winter entry, (2017), available at: https://en.wikipedia.org/wiki/AI_winter, accessed 22.08.2017

¹²⁵⁶ The limitations of deep learning, (2017), available at: https://blog.keras.io/the-limitations-of-deep-learning.html?utm_content=buffer3e94c&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer, accessed 22.08.2017

Technically the following may hinder development:

- Purported algorithmic successes of deep learning and neural network technology will reach a limit of what it's capable of achieving and will relegate it to a narrow technological domain whereby it can't attain a generalization capability that would allow a single algorithmic structure to permeate and support general problem solving capabilities and embody thought processes.
- Hardware components reach physical limits that simply can't be overcome

Technically the following may accelerate development:

- New computational models may allow for dramatically increased levels of computational ability e.g. HP's 'The Machine' architecture¹²⁵⁷, quantum computing methods
- An algorithmic breakthrough that allows neural network technology to achieve generalization capabilities, meaning that a problem solved in one domain can be used cross-contextually
- Wide-spread adoption and acceptance of AI-powered assistant agents

Societal factors that may hinder development:

- AI hacking: Malicious users attack AI models, manipulating and/or destroying them compromising trust, integrity and validity
- Users groups resist automatic decision making processes embedded AI brings to the technological space, resulting in a back-lash against adoption
- General fears that arise from 'Terminator'esque' scenarios being played out
- Accountability to an Artificial entity causes psychological repercussions towards humanity, as blame assignment seems unsatisfactory

2. Is the current level of excitement which surrounds artificial intelligence warranted?

Yes and No. We are clearly experiencing, what is known in the technological industry as a 'hype-cycle' and AI along with cyber security are two avenues that are peaking. At the outset, it may appear that the technologies propelling the changes are offering little new on the horizon, however this must be caveated with a word of caution.

Despite many in the field jumping on the proverbial AI hype-cycle, key innovators are achieving things that many thought would not be possible in their lifetimes and for this reason, commentators, observers, governments and policy makers need to pay attention.

AI is offering a lot in terms of automation and contextual prediction. When you couple the developments in this field with a 'connected-society', increased deployment of internet-enabled devices (Internet of Things) and ability to amass, store and analyse huge data-sets, a great deal of transformative possibilities become available.

If AI's purpose is to create computers that think and perform like humans, caveats must be made that what is being created has some bearing on human

¹²⁵⁷ HPE: The Machine, (2017), available at: <https://www.labs.hpe.com/the-machine>, accessed 22.08.2017

behavior, however it's still a grounded computational agent. There is plenty of reason to be excited about what is coming, however if you're rooting for 'strong AI', you'll probably be left wanting. 'Weak AI' on the other hand looks to be getting more advanced on a daily basis.

23 August 2017

Jonathan Sinclair – Supplementary written evidence (AIC0035)

House of Lords: Select Committee on Artificial Intelligence

Call for Evidence

Question 9: In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

9. 1 In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable?

Artificial intelligent (AI) systems and the back-end algorithms that are powering the surge in recent developments: neural networks and deep learning architectures, will start/are starting to be embedded ubiquitously in technology, from finance to automobiles, to entertainment.

When the systems are operating as expected the performance increases are dramatic however when things go wrong, accountability is left floundering and, as demonstrated in the 1980's with the Pentagon's attempt to automatically identify tanks¹²⁵⁸ and more recent investigations highlighting the brittle nature of these technologies¹²⁵⁹, the acceptance of ignorance concerning how these technologies operate is coming under increased scrutiny where even chat bot behavior is being deemed unacceptable by society at large.¹²⁶⁰

This back-lash from society makes it difficult to determine where the acceptance of black-box AI systems is permissible and where it should be reined in, held accountable and actions exposed through recently emerging mechanisms, colloquially termed 'explainable AI',¹²⁶¹ having said this, some determination must be made as to when transparency is permissible and when it is not.

It seems that society is the gatekeeper for setting the limits however it is the authors opinion that AI's lack of transparency is acceptable in the following areas:

¹²⁵⁸ Neural Network Follies, (1998), available at: <https://neil.fraser.name/writing/tank/>, accessed 29.08.2017

¹²⁵⁹ Machine Learning is Fun Part 8: How to Intentionally Trick Neural Networks, (2017), available at: <https://medium.com/@ageitgey/machine-learning-is-fun-part-8-how-to-intentionally-trick-neural-networks-b55da32b7196>, accessed 29.08.2017

¹²⁶⁰ Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism, (2016), available at: <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>, accessed 29.08.2017

¹²⁶¹ Wikipedia entry: Explainable AI, (2017), available at: https://en.wikipedia.org/wiki/Explainable_AI, accessed 29.08.2017

- Automated preference selection e.g. with online retailers through distributors like Amazon, High-street retailers, online-music distribution networks etc.
- Automated vehicle routing and re-routing when decoupled from an actual device e.g. in GPS where the user can influence choice
- Weather prediction
- Computer games
- Etc.

It shouldn't be a requirement where user choice is still an option and the resulting decision doesn't result in the possibility of harm occurring to another living entity.

9.2 When should it not be permissible?

AI transparency should be a requirement in all other cases not included in the aforementioned definition i.e. where the AI system presents a predication where no choice is presented and harm to another living entity may occur e.g. the determination of a medical diagnosis, the identification of a threat, the decision to avoid an obstacle within the context of an automated vehicle, the determination of a legal outcome, automated weaponry, etc.

30 August 2017

SiteFocus Incorporated – Written evidence (AIC0187)

The pace of technological change

Question 1: What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

SiteFocus Response to Question 1:

1. The current state of Artificial Intelligence (hereafter, AI) is being driven by Big Data and vast computing resources that enable the exploration of various statistical models. Predictive analytics enables AI algorithms to draw reference from large sets of historical data to depict the probabilistic outcome of current events.
2. Prevailing AI is limited to what is stated (i.e. unless it is learned from past data, the current state of AI cannot manifest any inference from new ideas, subjects, perceptions, or geopolitical events). Moreover, the reliance on historical data has created a prerequisite for large datasets which has resulted in the limited application of AI.
3. The AI of Tomorrow will evolve into AGI – Artificial General Intelligence. AGI requires a new way of thinking in the research and development of machine intelligence. Statistical approaches must be reinforced with the application of deductive and abductive reasoning to accomplish AGI. In this regard, SiteFocus is moving towards this goal with our pioneering work on the CIF platform.

Question 2: Is the current level of excitement which surrounds artificial intelligence warranted?

SiteFocus Response to Question 2:

1. In a word, yes. Today's AI is redefining automation as previously known and creating unprecedented benefits to human productivity. These benefits are manifested in applications such as autonomous cars, robotics, language translation, voice commands in popular devices (i.e. Amazon Alexa, Apple's Siri, Google Voice) which enable consumers to enjoy greater conveniences across different spaces. The deployment of AI in pattern recognition of images and medical applications has and will continue to create huge benefits to mankind.

Impact on society

Question 3: How can the general public best be prepared for more widespread use of artificial intelligence?

SiteFocus Response to Question 3:

1. The general public will largely be unaware of the deployment of artificial intelligence in everyday life. Rather, they will enjoy tangible and intangible benefits of its application. Such benefits will manifest as conveniences across a myriad of channels that touch everyday products and services such as transportation, security, access to products and services.

Question 4: Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

SiteFocus Response to Question 4:

1. Enterprise who can deploy artificial intelligence for horizontal application that benefit business and consumers alike stand to gain the most economic value. Consumers stand to enjoy the greatest benefits of AI-enhanced products and services.
2. People who do not use technology in the course of everyday life stand to gain the least from the development of AI. Government plays a significant role in offering the benefits of AI to its citizens through public services (i.e. connected cities via public transportation, infrastructure, energy, security).

Public Perception

Question 5: Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

SiteFocus Response to Question 5:

1. It is important to educate the public on the definition of AI (in its current state), its actual abilities, current implementations and, more importantly, the limitations of such implementations.
2. Government must play a huge role in educating the public. Particular effort must be made to demystify AI. This is key to increasing user adoption and avoiding "AI-phobia" or the creation of unrealistic expectations arising from misinformation, which can lead to catastrophic consequences.

Industry

Question 6: What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

SiteFocus Response to Question 6:

1. All sectors stand to benefit from the use of Artificial Intelligence. The depth of benefit will vary based on the sector. Sectors that emphasize and/or center around creativity (i.e. humanities, the arts, literature, music) stand to gain the least from AI. Creativity is the antithesis of AI.

Question 7: How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

SiteFocus Response to Question 7:

1. Data is created when a user engages an application provisioned by a business or government organization. The democratization of data collection should be safe-guarded by law and regulation to preserve individual rights and privacy.
2. While organizations with advantages derived from data may manifest an initial edge over its competitors, the limitations of Big Data analytics will render such data-driven advantages less effective. Big Data can be viewed as a silo. Its source is limited to the extent the organization has means to collect data. As a result, any implementation of that data is limited and can easily lead to false assumptions. A broader, democratized view of data derived from the world is much more important. Data is dynamic. Democratized data enables a broader view unconstrained by the limitations of data collection methodologies.

Ethics

Question 8: What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

SiteFocus Response to Question 8:

1. In its current state, AI is unable to provide a detailed accounting of its data derivatives and algorithms. Without knowing the inner-workings of an AI solution, users and organizations are left must treat it as a "black box".
2. The lack of transparency may create an environment where liability for unintended consequences can be deferred, absolving users and/or organizations from the responsibility of decisions made or actions taken,

based in whole or in part, on the AI. For example, a doctor could order a computed tomography (CAT scan) of a patient and, with guidance from an AI solution, incorrectly diagnose – or worse, overlook – a cancerous tumor because the black box algorithm made a mistake or experienced an exception.

3. As discussed in Response to Question 9, AI solutions may also foster a sense of overconfidence and/or reliance that may result in catastrophic consequences. Using a limited dataset to create larger conclusions which exceed the bounds of the data can have inconsequential (i.e. an unfavorable movie recommendation) or catastrophic consequences (i.e. an autonomous driving accident resulting from the AI confusing the broad side of a white tractor trailer with a clear, open sky).

Question 9: In what situations is a relative lack of transparency in artificial intelligence systems (so- called 'black boxing') acceptable? When should it not be permissible?

SiteFocus Response to Question 9:

1. Black boxing should not be permissible in (1) life-threatening situations or (2) decisions (based on AI) that may lead to irreparable harm. These guiding principles should form the basis for evaluation of AI adoption and deployment.

The role of the Government

Question 10: What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

SiteFocus Response to Question 10:

1. In the United States, there are existing government agencies that regulate energy, environment, and transportation. Similar agencies exist in governments across the world. With the adaptation of AI, our response to Question 9 should be used as a guiding principle for passage of law and regulation on such AI enablement.

Learning from others

Question 11: What lessons can be learnt from other countries or international organisations (i.e. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

SiteFocus Response to Question 11:

1. AI implementations are still relatively new. However, several notably public examples have proven that unregulated AI deployments can be

problematic to the public safety. These examples include, but are not limited to common facial recognition errors in law enforcement engagements and high-frequency trading in the U.S. financial services sector (i.e. the Flash Crash of May 6, 2010).

Prepared by: Cameron K.F. Koo

For: The House of Lords, Artificial Intelligence Committee

Dated: 2017-09-06

6 September 2017

Dr Will Slocombe – Written evidence (AIC0056)

EVIDENCE FOR HOUSE OF LORDS SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE

Dr Will Slocombe, 3 September 2017

1. This evidence, submitted in an individual capacity, is based upon my role as a lecturer in the Department of English at the University of Liverpool, undertaking research into representations of Artificial Intelligence, and drawing upon my work in engaging the public with such representations over the last three years. It is related to my individual written evidence to the Robotics and artificial intelligence inquiry of the Commons Science and Technology Committee in 2016 (ROB0015), although it covers slightly different areas, based on the questions in the Call for Evidence. As with that earlier evidence, I am not specifying particular representations within this statement for the sake of brevity, but rather dealing with general cases and principles.
2. This statement focuses on a limited number of the areas addressed by the Committee's Call, concerning the pace of technological change, impact on society, and public perception, and primarily upon the role that fictional representations of AI play in public understandings of the technology. It is to be considered evidence of the ways in which expectations and fears about AI have been often framed by a set of assumptions generated by dominant representations, and thus how these representations are serving to inform and/or misinform the public about the potential benefits and dangers of the technology.

Broad overview of the pace and direction of technological change

3. According to the majority of representations of AI, particularly those produced within the last twenty years, the possibility of Artificial General Intelligence (AGI) remains a distant possibility, as the settings of representations dealing with high-level AGI systems are often in the far future. Whilst there is a relative balance in the representations between AI as being of social utility and as being of great existential danger, cinematic representations tend towards the more pessimistic whereas literary representations tend to be more nuanced in their approach to the topic.
4. Broadly speaking, representations set in the far future do not engage with the pace of technological change concurrent with the development of AI other than suggesting that there is an acceleration in technological progress preceding the development of AGI, an immediate expansion of applications (both positive and negative), followed by a period of significant social upheaval as the true ramifications of the technology

become apparent. Such upheavals tend to be economic, and lead to a significant reappraisal of social roles and the types of work that humans can perform even at when not apocalyptic and at their most positive.

5. Representations set in the near to middle term tend to focus on specific systems and applications (drones, data mining, robots designed for specific tasks), rather than human-level intelligence. These representations suggest that AI systems will be utilised across a variety of industries, but that they will be “owned” by the individuals and companies that produce them, and such companies are therefore the primary drivers behind the technology.
6. In both of the above cases, commerce and industry are assumed to drive the development of AI, more than academic or private efforts, and within those two economic spheres, it is primarily technology companies dealing with distributed systems (AI software programs) rather than “embodied” AI (industrial robotic systems) that are discussed. Developments tend to emerge within an unregulated, market-driven environment, but these are often responsible for unforeseen (and negative) consequences as they are brought “online” without considering the larger social impact that they will have.
7. The fears that the main representations of AI embody are often therefore extensions of fears about corporate control and regulation (who has access to data and what can they do with it? who “controls” the technology? what happens if the AI are autonomous?). In such ways, there is an underlying concern at the extent to which society is being “mechanised” that is often being espoused, and such representations evince a concern for a loss of humanity, individuality, and choice. Very few representations of AI are concerned with the technology’s potential at all, but instead serve as a means to think through different areas of social concern and human autonomy.

Impact on society & public perceptions

8. It is through fictional representations and the media that the public engage with the advantages and disadvantages of AI technologies. With the media often relying on dominant tropes from fictional sources, however, it is the fictional representations that become the primary metaphors through which the technology is perceived, despite the fact that they are fictional. Paragraphs 3-7 outline the primary representations of AI that currently exist and it must be realised that these therefore become the dominant representations of AI with which the public are familiar. Although the representations and their tropes are often acknowledged to be “science fiction” it is nonetheless the case that science

fiction provides “shortcuts” for thinking about AI, however questionable their legitimacy.

9. As a result of such dominant representations, the “Terminator / Skynet” metaphor for AGI or the “Three Laws” understanding of programming parameters and ethical and legal responsibility come to stand for the truth of developments in AI technology, whether they are plausible or not, and despite the fact that only a limited number of examples tend to be drawn upon. That is, although science fiction and speculative fiction are often dismissed for their lack of “truth”, they nonetheless inform social expectations and understanding of technologies such as Artificial Intelligence, as well as to a lesser extent—through programmers’ and developers’ engagement with such representations—impact upon the ways they might eventually manifest. More importantly, discussions about AI tend to resort to a few representations, and generally not the more positive ones, meaning that there is a lack of balance to the available representations that are most often discussed.
10. Through various outreach and engagement events, I have worked on engaging the public with fictional representations of Artificial Intelligence in order to encourage a more informed consideration of the issues the technologies and their representations raise, and a more sustained engagement with the breadth of such representations. The most useful format has been working with an existing representation or series of representations (such as by screening a popular film or reading a famous text or set of short stories) and then encouraging the audience to consider how such representations function, in relation to either particular implementations (such as robot carers) or broad issues around AI (such as machine consciousness). An interdisciplinary panel discussion following such a screening or reading is then used to draw out the factual and fictional elements of the representation and, through subsequent wider discussion, enable participants to consider their own views on the subject.
11. For instance, in one such engagement event the panel consisted of a computer scientist working on verification and validation, a philosopher working on philosophy of mind, and a literary critic working on representations of AI; after a film screening, each panellist presented a short reading of the film from their own disciplinary perspective, followed by break-out sessions led by the panellists. As a result of this, the audience seemed to engage with several important ideas concerning the ways in which that representation of AI functioned: participants engaged with the politics and artistic effects behind the representation itself, considering the ways in which particular elements of the film were fictionalised for the purposes of dramatic effect whereas other elements were more accurate to current research in the field and, moreover, had the opportunity to articulate what they considered the main issues around

AI to be, as well as related areas such as data management and the ubiquity of social media. This meant that rather than a particular representation of AI being integrated into the audience's view of the technology, adopted as yet another metaphor to be mapped onto participants' expectations, the event demonstrated the value of humanities and STEM disciplines by using various methodologies to consider what was fictional versus what was factual, the difference between dramatic effects and real-life programming parameters, and furthermore ask what elements of the technology seemed useful or worrying to members of the audience.

Summary

12. I firmly believe that (fictional) representations of Artificial Intelligence can serve to inform the public about AI technologies precisely because they already do, and it is through engaging with such fictional representations that discussions about what is possible, what is necessary, and what is desired can be brought to the fore. It seems to me that, rather than being dismissed as merely "fictional", it is precisely because of their very fictional natures that speculative and science fictions can be used to engage the wider public in debates about AI technologies, and used to demonstrate both the limitations and opportunities of the technologies and the representations themselves.
13. In this sense, fictional representations can be used to stimulate public discussion and consideration of the current, as well as possible, state of AI research and implementation in relation to the broad gamut of issues that the Committee is addressing. For each question that is asked, there is a likely an existing or forthcoming representation of AI that considers that problem, and using the representation to provoke different answers to the question, and different scenarios to be considered, can be of great benefit to wider engagement with the actual realisation of the potential of the technologies.
14. Although there is precedent in the application of science fiction films to discussions about a new technology or the legal framing of scientific advances—such as the relationship between the science fiction film *Gattaca* (1997) and debate in the US about genetic determinism and the Genetic Information Nondiscrimination Act (2008)—a more structured and considered approach would be required in relation to AI. I would propose that the most effective way of engaging the public with the topic, and ensuring that it is suitably well-informed, would be to run a series of specific interdisciplinary workshops alongside various sectors developing AI tools, public organisations, and different sections of the public (particularly those likely to be most disadvantaged by the wider implementation of AI technologies). These workshops would, using

selected representations of AI, be designed around the main issues that the Committee wishes to address. Using such representations as a springboard into public dialogue and debate ensures that the representations can be viewed critically, that the dramatic tensions they rely upon can be talked through from various perspectives, and that the most recent developments and trends in the research in the area can be presented and discussed openly. Most importantly, they should not be a set of presentations or talks, but very deliberately shaped to be participatory and discursive so as to generate the most discussion and feedback, and therefore help the public to shape the future of such technologies rather than being passive bystanders in their development.

3 September 2017

Dr Chris Steed – Written evidence (AIC0017)

Future Proofing?

A proposal to address the 'future shock' arising from artificial intelligence

Submission to House of Lords Select Committee on Artificial Intelligence Dr Chris Steed FRSA, EeD (Exeter), Arches Project Totton

"The illiterate of the 21st century will not be those who cannot read or write; but those who cannot learn, unlearn and re-learn" – Alvin Toffler

Another world is happening. The world is spinning faster and faster; so fast in fact that we are educating children for a society that will be out of date within fifteen years.

The Select Committee on Artificial Intelligence appointed by the House of Lords on 29 June 2017 is considering the economic, ethical and social implications of advances in artificial intelligence. This is of critical importance and high urgency. As Lord Clement-Jones, Chairman of Committee has said, *"This inquiry comes at a time when artificial intelligence is increasingly seizing the attention of industry, policymakers and the general public."*

The Select Committee is endeavoring to understand what opportunities exist for society in the development and use of artificial intelligence, as well as what risks there might be. Our focus here is on an important dimension to how we might gear up as a country to a digital era.

Future Shock

The world of work is changing faster and more drastically than at perhaps any other time in recent history. According to research from the World Economic Forum, 35% of the skills necessary to thrive in a job today will be different five years from now. My little grandchildren can expect to change jobs at least seven times over the course of their lives – and five of those jobs don't exist yet.

This revolution is arriving on the back of a slew of transformative technologies. But it is much more than the sum of these technologies. The first industrial revolution came in on the back of a wave of innovation – the invention of the steam engine and the cotton mill, for instance – and represented a history-altering wave of systemic change such as urbanization, mass education and industrialization of agriculture. The second industrial revolution, with electrification and mass production, saw the advent of entirely new social models and ways of working, and the third industrial revolution – the digital revolution – provided the electronic and computing foundations for the radical shrinking of the

world we have seen over the past five decades. The Fourth Industrial Revolution is fundamentally changing the way we live, work and relate to one another. You've seen nothing yet!

I observed at a St Paul's Institute on 'the Future of Work' recently ¹²⁶² that work is not just what allows us to sustain ourselves and our families: it is our primary source of validation and value, it expands our possibilities to be and do what we aspire to; it gives meaning and rhythm to our endeavours as part of a broader social contract.¹²⁶³

Yet it is a massive undertaking to prepare students adequately for a world of work when the landscape will be all but unrecognisable by the time they get there. Price Waterhouse recently produced a Report estimating that almost one third of UK jobs by the 2030s could be affected by advances in automation. John Hawksworth, their chief economist, said that the jobs most at risk are those that are "more manual, routine jobs" which "can effectively be programmed." "Jobs where you've got more of a human touch, like health and education," would be less affected, he said. New automation technologies will both create some totally new jobs in digital technology. Such productivity gains will generate additional wealth and spending that will in turn undergird additional jobs in services sectors that are not less susceptible to automation.

Future proofing

How can we prepare for a workplace of the future if we're not quite sure what it will look like? What skills or expertise should students focus on acquiring today if they want to succeed tomorrow? The answer seems to be technical skills such as coding, and all that makes for digital intelligence. But soft skills – such as teamwork are going to be really vital. As machines push us to specialize in our competitive advantages: more "human" work, creative and social intelligence, interpersonal and non-routine tasks are what makes us resilient and adaptive to change. This too is what makes us human. The jobs that even artificial intelligence can't replace will be those that require strong human character traits. Rather than technologies for their own sake, workers will need empathy – the ability to persuade and to work well with others and creativity to apply existing knowledge to the wide-reaching changes to business, society and politics.

As the head of technology and investments at PwC, Jon Andrews, observed, "in the future, knowledge will be a commodity so we need to shift our thinking on

¹²⁶² Future of Work- St Paul's Institute 28th November 2016

¹²⁶³ Leonardo Quattrucci Policy Assistant to the Head of the European Political Strategy Centre, European Commission WEF 25th November 2016

how we skill and upskill future generations. *Creative and critical thinking will be highly valued, as will emotional intelligence.*"¹²⁶⁴

Brynjolfsson and Mcfee ponder whether human workers can upgrade their skills fast enough to compete with rapid automation. In the global society that is already arriving, creativity and empathy may be given the same status as numeracy and literacy because learning to collaborate and learn (or unlearn) are skills the future will value. These are gifts of the imagination less easily automated. ¹²⁶⁵

At Matthew Taylor's suggestion, I wrote these words in a blog on 17th March for the website of the Royal Society of Arts.

"Curriculum reform for a digital era seems to have everyone's attention these days. In March, eleven university presidents from South Korea, Indonesia and Turkey tackled the issue during a roundtable meeting at the *Times Higher Education Asia Universities Summit*, hosted by South Korea's University of Ulsan and Ulsan Metropolitan City. 'I think all the universities all over the world are forced to reform their curricula to cope with the coming new era,' said Gu-Wuck Bu, president of South Korea's Youngsan University. But he added: 'The real problem is the future is unpredictable. We cannot say what kind of job will disappear and what will survive.'¹²⁶⁶" A Global Education and Skills Forum (GESF) in Dubai in March 2017 discussed the notion of a Digital Quotient. In the same way as IQ and EQ measure general and emotional intelligence, DQ measures a person's ability and command of digital media. The DQ Institute observed that DQ was identified by the World Economic Forum as an effective way of improving digital citizenship."

A recent book ('Smart Leadership, Wise Leadership: environments of value in an emerging future') ¹²⁶⁷ suggested that educating for a digital future should address these concerns so as to ensure that a 'humans only' zone of life and work is both retained and sustained. Data is one of the driving forces of the Fourth Industrial Revolution. But sometimes, when we perceive the world through data-driven models, it becomes harder to see the humanity behind the numbers. Technology thus has the potential to erode our sense of empathy.

An inter-generational pilot project

¹²⁶⁴ Market Business News March 24th 2017 ref 'Consumer Spending Prospects and the impact of automation on jobs UK Economic Outlook March 2017

¹²⁶⁵ Brynjolfsson, E. & McFee, A. (2014) *The Second Machine Age: Work, Progress and Prosperity in a time of Brilliant Technologies*. New York: W.W.Norton

¹²⁶⁶ john.morqan@tesglobal.com accessed 10th March 2017

¹²⁶⁷ Steed, C. D. (2017) *Smart Leadership, Wise Leadership: environments of value in an emerging future* London: Routledge

On the edge of Southampton, we're developing a pilot project inviting area secondary and further education providers to become linked to an all-age creative and empathy hub.

Within the Diocese of Winchester, a social enterprise company has been formed (called 'the Arches') arising from an initiative to regenerate a significant space, the largest of its kind in Totton, as a hub for inter-generational community engagement to help transform the lives of young and old, especially the socially isolated. Under these plans, there is considerable scope to develop a project with education providers such as our nearest further education vocational skills college and a large comprehensive in order to enhance employability for a digital future.

The development of a creative arts environment as a rich context for addressing social issues has come to the fore, especially the growing problem of isolation in our communities that arises from a fragmented society. We have had contact both with the Jo Cox cross - Parliamentary Commission on Loneliness and also the recent All- Party Group on Arts, Health and Wellbeing about this. Area health people and local GP practices within the West Hampshire CCG are solidly behind what we are doing with the aim of keeping people active and creative through constructive tasks. The model is that of co-production of efforts to sustain wellbeing and crucially, social prescribing to the Arches project. It will sit alongside such schemes as Timebank and befriending to promote renewal of community bonds.

In tackling social isolation through the arts and music, and thus potentially saving the system a considerable amount through the social value it generates, the pilot project is being designed to be inter-generational. We hope to include a dimension of a recently publicised approach in which all-age nursery provision brings younger and older members of the community together very fruitfully. Crucially, it will feature learning opportunities that go both ways – such as young people mentoring older members of the community in IT and in return, receiving life wisdom in the context of CV writing and employability from experienced people. Within a creative eco-system, there is huge potential to learn aptitudes and new patterns of creativity and empathy that digital futures require.

The prize to be gained from such inter-generational exchanges is the nurture of empathy, collaboration and insights into team work. Ultimately the benchmarks could be assessed and publicly accredited in ways that future employers would respect.

Creativity – creativity might be thought as the next barrier to fall as robotics learn to script texts and perform creative tasks. The difference is that though there is pattern recognition, there is no meaning attached to the symbols. They are not signifiers of anything. It is the human dimension that brings true creativity because it comes from and generates meaning. Creativity is not just

about innovation and the capacity to think in new ways, it can arouse empathy by the route of imagination. Eyes that now see may, under the right circumstances, lead to a mind and heart that are now open. It is well-known that creativity guru Sir Ken Robinson challenges profoundly the way we are educating our children.¹²⁶⁸ He is a vociferous champion of a radical re-think of our school systems, to cultivate creativity. We have been educated to become good workers rather than creative thinkers. Students with restless minds and bodies – rather than being hailed for their energy and curiosity are ignored or even stigmatised with costs and consequences. This has implications for leadership development – something we are aiming to address in an innovative way through the Totton project.

Empathy - is the skill of grasping what others think, feel and perceive. It is not only active listening, but the practice of imagining or inferring other minds. Data literacy (to make sense of the torrents of information that will continue to emerge) is clearly vital but so is adaptability and the emotional intelligence to apply it to people we need to deal with as we face huge fresh challenges. A strong dose of empathy will be vital. Perhaps the great challenge of the next 10 years for corporations and institutions will be to rebuild the empathy that we've lost.

In addition to community schools and places of worship, there are many heritage spaces around that make ideal social eco-systems of this sort. For decades, the emphasis in education has been about imparting knowledge. Offering experiences to learn creativity and empathy is a vital combination of social skills increasingly needed in the digital economy. To ensure there are sufficient jobs of quality and not just quantity, education systems need re-tooling. As a report by an economic think tank, the Hamilton Project, argued, it is '*Goodbye, maths and English. Hello, teamwork and communication!*'¹²⁶⁹

The Hamilton Project identified four trends in the workplace that are relevant to this:

- 1. Today's jobs demand more non-cognitive skills than they did in the past.**
- 2. The labour market increasingly rewards non-cognitive skills**
- 3. Students who develop them are more likely they are to be in full time employment.**

¹²⁶⁸ Robinson, K. (2006) *The Element: How finding your passion changes everything*. NY: Barnes and Noble. See also his *Out of Our Minds: Learning to be Creative*. NY: Wiley/Capstone

¹²⁶⁹ Whitmore, D. et al Hamilton Project Seven Facts on Non-cognitive Skills from Education to the Labour Market www.hamiltonproject.org economic facts October 2016

4. Those with fewer non-cognitive skills are being left behind.

Tasks involving working with or for people – requiring non-cognitive skills - are substantially more important now than they were in the 1980s and 1990s. The need for social skills and service skills grew by 16 and 17% respectively, while tasks needing high levels of maths have only grown by 5%. As non-cognitive and cognitive skills rise, so do earnings and probability of full-time employment.

It is to generate a creative environment where people are relating at the level of concerns and interests is one where they are continually learning together, expanding their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured and where collective aspiration is set free. This is where creativity and empathy education come together very fruitfully as key skills to future-proof our children and growing future leaders. They are far less susceptible to technological displacement. Moreover, creativity and empathy education are central to positive cultures that foster the value of the human. As Albert Einstein remarked, *"I do not teach my pupils. I only create the conditions in which they can learn"*.

Conclusions

It is certainly far from easy to grasp the changes that we are seeing around us. In something I have written for Routledge, it seemed fruitful to try to analyse the puzzling times through a particular lens, that of the 'death of distance' as technology shrinks the planet but also the attitude of 'keep your distance' as people build walls. This is of a piece with the response to de-personalisation we have seen across the board in recent times.¹²⁷⁰

I refer to how in the run up to the American election, globalisation was blamed for taking American jobs. The real culprit there was probably automation. It was robotics that are stealing blue collar jobs rather than Asian workers – witness the petrochemical plants in the USA that don't need working class workers. A PhD in chemistry plus robots will do the job more cheaply. How President Trump will overcome those structural forces is far from clear.

The World Economic Forum Future of Jobs report argues that emotional intelligence, creativity, and people management will be among the top skills needed for jobs in 2020. "Change won't wait for us: business leaders, educators and governments all need to be proactive in up-skilling and retraining people so everyone can benefit from the Fourth Industrial Revolution," the report states.¹²⁷¹

¹²⁷⁰ Steed, C. (2017) *'We Count, We Matter: Voice, Choice and the Death of Distance'* – Routledge (forthcoming)

¹²⁷¹ The Future of Jobs Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution Global Challenge Insight Report World Economic Forum January 2016

There is growing awareness that technological displacement will radically re-shape the workplace within the next twenty years. After all, if a routine task can be performed cheaper, faster and better by a robot, there is a chance it will be. A premium will be placed on factors to do with the human 'touch' such as creativity, empathy and entrepreneurial flair that cannot be replicated by algorithms. How we re-humanise the workplace and social ecology so as to generate non-economic 'human value' will be an increasing challenge of our times.

In the global society that is already arriving, numeracy and literacy will not be the only skills that State education systems prize to help them compete. Creativity and empathy may be given the same status as numeracy and literacy because learning to collaborate and learn (or unlearn) will be soft skills the future will value. It is high time to make investing in these soft skills an education priority and have a national conversation about it.

17 August 2017

Professor Richard Susskind – Written evidence (AIC0194)

Artificial Intelligence – Challenges for Policymakers Professor Richard Susskind OBE FRSE

*A Submission to the House of Lords Select Committee on Artificial Intelligence
6 September 2017*

1. In this note, I identify and discuss five pressing policy challenges that will arise from the widespread use of artificial intelligence (AI) in our working and social lives.
2. My background is in law and technology, my doctorate in the mid-1980s was in AI and law, I have written several books on this subject, and I co-developed in 1988 the world's first commercial AI system in law. I am President of the Society for Computers and Law, Chair of the Advisory Board of the Oxford Internet Institute, and Strategy & Technology Adviser to the Lord Chief Justice. I write here, however, in a personal capacity.
3. Much that follows draws from the arguments and findings presented in *The Future of the Professions* (OUP, 2015; paperback 2017), a book that I wrote with Dr Daniel Susskind, a Fellow in Economics at Balliol College, Oxford. If the Select Committee seeks greater depth of analysis, especially on the impact of AI on white-collar work, that book, including its bibliography, may be a useful further source.
4. Consistent with the categories suggested in the call for evidence, this submission is organized as follows: after some preliminary remarks about AI, I present a working hypothesis about the broad future direction of AI, and then discuss the following five topics and associated challenges:
 - pace of technological change – long-term planning;
 - impact on society – employment and education;
 - industry – competitive strategy;
 - ethics – the limits and ownership of AI; and
 - role of government – implications of Brexit.

Preliminary

5. I have always found the term 'AI' both helpful and unhelpful. The upside is that the concept often generates curiosity and excitement and, in turn, the field frequently attracts first-rate entrepreneurs and technologists as well as substantial investment. The downside is that no-one seems entirely clear what the term means and it is often wielded as no more than a rather blunt marketing weapon or as part of an alerting headline or tweet.

6. There are two broad ways to define AI. The first is 'architectural', in terms of the tools and techniques used. When I worked on AI in the 1980s, the technological fashion was for rule-based systems and logic programming. This was the first wave of AI – systems that were explicitly programmed to undertake tasks by, essentially, following huge decision trees and flowcharts put together by human developers. Today, different methods, like supervised machine learning and deep neural networks, are very popular. This is the second wave of AI – instead of following explicitly articulated rules, these systems learn from large bodies of past data.
7. Ordinarily, however, technical terms and concepts mean little to most non-specialists, for whom a second type of definition – 'functional' – is more useful. When we speak about AI in functional terms, we are talking about what these systems actually do, what tasks they undertake. And, very generally, when many AI specialists and others refer today to AI, they are speaking of systems that perform tasks (cognitive, creative, manual, and even emotional) that in the past we thought required the thought processes of human beings. This remains a loose characterisation of AI but what is significant on this account is that machines are performing more and more such tasks. Our systems and machines, as Daniel Susskind and I say, are becoming *increasingly capable* – this phenomenon, rather than specific techniques and technologies, should be our main focus when exploring the economic, ethical, and social implications of AI.

Working Hypothesis

8. In this submission, I assume the following line of argument (this is a simplification of the position laid out in *The Future of the Professions*). Leaving the term 'AI' to one side for now, it is clear that (a) our systems and machines, as noted, are becoming increasingly capable; (b) they are taking on more and more tasks that were once the exclusive province of human beings; (c) although new tasks will doubtless arise in years to come; (d) machines are likely in time to take on many of these as well.
9. Many people respond that there are limits to what machines can do. They accept systems can undertake 'routine' work but contend that there are many 'non-routine' tasks – creative and emotional ones, for instance – that only human beings can perform. They challenge (d) above and insist that when traditional jobs fade, new ones will always emerge, made up of those tasks that are beyond the reach of even the most capable machines. However, our extensive research into professional work does not support the view that the new tasks that emerge are and will be ones for which humans are better suited than machines. It transpires that insistence that there are tasks that can never be undertaken by machines often rests on what Daniel Susskind and I call the 'AI fallacy' – the belief that the only

way to develop machines that can perform at the level of human beings is to *copy* the way human beings work. The error here is to fail to notice that many contemporary AI systems operate not by copying human beings; instead, they function in quite different and *unhuman* ways.

Pace of technological change

10. My current sense is that much of the current debate on the impact of AI overstates its short-term impact and that we are approaching the peak of the 'hype cycle'. Bill Gates once said, to paraphrase, that less happens in two years than we expect when it comes to technology, but more happens in ten. So too with AI. I do not anticipate fundamental societal change in the next 18-24 months. However, I believe that the long-term effects are invariably understated. I anticipate that, in the mid to late 20s, the impact of AI on our personal lives and our social, political and economic institutions will be pervasive, transformational, and irreversible.
11. Aside from the substantive upheaval that AI will bring about, this projection for the 2020s poses a major challenge for policy-makers and strategists alike. In most public bodies and businesses, long-range planning rarely extends beyond five years or so. Given the likely impact of AI in the time scales I anticipate, businesses should greatly lengthen their strategic planning time frames, while the government somehow has to look beyond the five-year political cycle that currently constrains visionary political thinking and investment for the long run.
12. In thinking about the policy implications of AI, the notion of increasing capability is important. There is no finishing line in AI. Nor are we plateauing. Every day we hear of some new development, system, app, or technological breakthrough. When discussing the future of AI today, all that most of us can do is extrapolate from what we already have. But we should acknowledge that it is likely if not probable by, say, 2025, that our lives will have been transformed by technologies that have not yet been invented. Accordingly, we should not assume that the leading enabling techniques of today (for example, machine learning) will dominate for the foreseeable future. There will no doubt be a third wave of AI; and a fourth; and so on. Policymakers should be both humbled and open-minded about *as-yet-uninvented technologies*.

Impact on society – employment and education

13. The hypothesis laid out above has significant implications for the traditional workforce. If machines are becoming increasingly capable, it is hard to avoid the conclusion that, in the very long run, much if not most human labour will be replaced. The best and the brightest workers will no doubt last the longest - those experts or highly skilled individuals who

perform the diminishing range of tasks that cannot or should not be replaced by machines. But there will not be enough of these tasks to keep armies of conventional workers in sufficiently paid work. I stress this is a very long-term view, by which I mean the 2030s or 2040s and beyond.

14. The medium-term position is quite different. Despite current claims in the media and by businesses that AI and robots are poised to replace all human jobs, it is more likely that the 2020s will be characterized by *redeployment rather than unemployment*. While many tasks, both of blue-collar and white-collar workers, will indeed be taken on by increasingly capable systems, the coming decade or so will be dominated by the development of these systems and by the corresponding need for new skills and disciplines, such as data science, machine learning, system development, and knowledge engineering. In the 2020s, accordingly, there will be new roles rather than no roles.
15. The call for new skills and disciplines should be made directly to our universities. One basic question must be asked and answered - *what are we training our young people to become?* My concern, in the professions at least, is that, despite the advances and trends noted here, we are nevertheless generating 20th century rather than 21st century graduates. What we teach and how we teach most of our aspiring professionals and white-collar workers has scarcely changed in the past 30 years. In truth, we are educating our young to undertake tasks (based on knowledge acquired by rote learning) for which machines will soon have comparative advantage over humans. We should want our graduates to excel in activities that are, for now, beyond the capabilities of machines. It would be misguided, however, to single out the universities as the sole weak link in building tomorrow's AI-based economy. A government-led wider review of our entire educational system should be undertaken, taking relentless advance in AI and the emergence of increasingly capable systems as its premise.
16. More challenging perhaps, we will also need to re-train much of our current work force if we in the UK, rather than others, are going to take on the new roles that will be important for economic success. If, as suggested below, we seek to lead the way in building these increasingly capable systems, we should want to involve people whose jobs have been replaced in system development. However, the gap between the current skill set of white-collar workers and the toolkit needed for the 2020s is large; and it is not always clear how this gap can actually be bridged. Beyond white-collar work, the position is yet more worrying – truck drivers who are rendered redundant by autonomous vehicles will rarely have the educational background or training to support their simple retraining and redeployment as, say, software engineers. A review of our educational system should also consider this question.

Industry - competitive strategy

17. In the 2020s, businesses and individuals will face the same simple and yet fundamental choice – *to compete with machines or build the machines*. For businesses, this is a question of strategy. For individuals, this is a matter of career direction. In either case, by seeking to compete with machines, this means a choice is being made to focus on doing things that machines cannot, even if that is a diminishing set of activities. By building the machines, I mean recognizing that there will be AI-based systems in the future that can undertake most work and tasks and so the preferred way to stay competitive is to be to be involved in actually building these AI systems.
18. Every British business should be addressing this strategic question – whether to compete with or build AI-based systems. The government too should be confronting this issue, as a matter of international competitive strategy and of domestic labour market policy. If it is accepted that machines are becoming increasingly capable and AI is thought to be advancing relentlessly, then one bold and radical proposition follows – that British businesses and British industry should seek *to lead the world in building the systems that will replace human workers*. If the displacement of many human workers by AI is projected, then there is a great opportunity to be a leader in the development of these systems. Conversely, looking at this defensively, if the UK chooses not to develop these systems, then other countries most certainly will. It is surely better that we develop the systems that disrupt our workforce and we export these systems to other countries than we have our workforce rendered redundant by systems that have to be imported.
19. Allowing that a call to build the systems that will replace human workers may be too radical a request, a weaker claim that is harder to reject is that economic prosperity in the 2020s will be enjoyed by those economies that pioneer in technologies such as AI and those that foster supporting R&D capacities and innovation programmes. In the UK, this calls for investment in our universities, in our start-ups, and in our mainstream businesses. Some of this funding will come from the markets but the state must also play a role, whether by offering tax breaks and seed funding or by taking advantage of any new flexibility outside the EU (for example, an immigration policy that makes it easier for the most talented technologists from beyond the EU to work here), and by cutting away the regulatory barbed wire that currently hinders so many companies.

Ethics – the limits and ownership of AI

20. Even if AI-based systems are in due course capable of taking on most or even all of the tasks currently performed by human beings, there may be some uses of technology that we would consider ethically unacceptable. Many would feel it wrong, for example, to leave it to a machine to pass a life sentence or turn off a life-support system. Some decisions, it can be argued, must be reflected upon and even agonised over by human beings. The buck cannot always stop at a robot, no matter how high-performing. I submit that a Government-led inquiry and public debate over the ethical boundaries of AI is urgently needed, so that we are clear and explicit about those tasks that we may never wish to be taken on by machines. I draw attention here to the analogy Daniel Susskind and I drew in *The Future of the Professions* with the debate in the UK, in the early 1980s, over the moral implications of emerging techniques such as in vitro fertilization and test tube babies. A national inquiry and consultation was launched, leading to an influential report by the philosopher, Mary Warnock. That inquiry generated great discussion amongst scientists, journalists, academics, and the public; and substantially raised the level of general understanding of the central issues. The main problems were clarified if not fully resolved at the time. Before our systems become much more capable, there is, I submit, a need for a similar scale of debate on the ethical limitations we should impose on AI.
21. Another set of vital ethical issues that require deep consideration relates to the ownership and control of tomorrow's AI, by which I mean the hardware (processing, storage, networks), software (algorithms, apps, packages) and data (personal, business, public) that in combination will play a central role in our economy and society. Already, we can see an unprecedented concentration of wealth and power in a small number of corporations (such as Apple, Facebook, Microsoft, and Amazon). One question is to what extent and how the activities of these businesses should be regulated. Another is to what extent and how the capital and revenues of these companies should be redistributed (as machines replace human labour, as capital becomes significant than wages, great economic and social inequality is likely to follow in the absence of state intervention). I recommend that the government-led inquiry into the ethics of AI should include the subject of ownership and control of AI.

Role of government - implications of Brexit

22. Britain's full disengagement from the EU is likely, in my estimation, to take a decade or so, whatever form our departure takes. This ten-year period will also be a time of greater technological progress than humanity has ever witnessed. Given the scale of interest and investment, advances in AI are likely to be especially rapid. However, if the UK is largely preoccupied

with Brexit, there is a danger that we miss the opportunity of emerging as a global leader in AI. If we become excessively introspective and neglect the advance of technology, our industries will struggle to compete. The US, China, Japan, and South Korea will not take a ten-year pause from investing and innovating in AI to allow the UK to keep apace.

23. Alternatively, we could regard the discontinuity of Brexit as an opportunity to rethink and rebuild Britain on the back of AI and other technologies. We could force ourselves to look at the post-EU world only through a digital lens. When there are new public institutions or businesses to be put in place, they should be AI-based in conception. We should lead the way and harness the power of technology in effecting Brexit and repositioning the UK.
24. In the public sector, this use of AI and other advanced technologies should transform and not simply streamline our current ways of working and governing.

6 September 2017

Professor Austin Tate, Professor Chris Williams, Professor Robert Fisher,
Professor Alan Bundy, Professor Simon King, Professor David Robertson and Dr
Michael Rovatsos – Written evidence (AIC0029)

**Professor Austin Tate, Professor Chris Williams,
Professor Robert Fisher, Professor Alan Bundy, Professor
Simon King, Professor David Robertson and Dr Michael
Rovatsos – Written evidence (AIC0029)**

[Submission to be found under Professor Robert Fisher](#)

techUK – Written evidence (AIC0203)

Introduction and Executive Summary

1. techUK welcomes the opportunity to provide written evidence to the Artificial Intelligence Select Committee inquiry. techUK is the industry voice of the UK tech sector, representing more than 950 companies who collectively employ over 800,000 people, about half of all tech jobs in the UK. These companies range from innovative start-ups to leading FTSE 100 companies. The majority of our members are small and medium sized businesses.
2. Our written evidence builds upon a number of the points raised previously in written evidence to the House of Commons Science & Technology Select Committee inquiry on the 'Robotics and Artificial Intelligence'. techUK would be pleased to provide the Committee with further information on any of the points made in the following submission.
3. Our top-level points for the Committee are as follows:
 - **Artificial Intelligence (AI) is a significant driver of change** across the UK economy and society. Used well AI is a power for good offering important social and economic gains to the UK – AI can boost productivity, economic growth and, if implemented and shaped correctly, personal and societal wellbeing.
 - **The UK has a position of strength in AI** due to a combination of factors. Including a strong digital ecosystem, a vibrant and competitive AI industry and world leading university and business R&D.
 - **AI is already enabling digital transformation** across sectors and industries including financial services, healthcare, transport, manufacturing and is a key driver in enabling digital entrepreneurialism. We are however only at the beginning of the development and adoption of AI technologies and more needs to be done to encourage investment and adoption.
 - **We must build consensus and greater confidence** on what the UK's AI driven future could look like and how we get there. This is an area where Government, industry and others must work together.
 - **We must also be vigilant to public concerns** that must be recognised and addressed. These include the impact on jobs, the privacy and security of data, whether AI systems are biased and profound social and ethics questions about the implications of a data driven future.
 - **A realistic, constructive and balanced discussion is needed on the impact of AI on the UK workforce.** This debate should not focus solely on how AI could replace people but consider how AI will create

higher-value added roles and could address the potential future reduction of the UK workforce.

- **A key challenge that remains however is the addressing the digital skills gap** in order to create a talent pool of skilled individuals as well as train, retrain and upskill the workforce in order to prepare for the future.
- **A culture of public trust and confidence in AI must be built** to ensure public support for increased development, adoption and use of AI technologies.
- **Mechanisms must be put in place now to discuss ethical questions being raised** and ensure technological innovation is on the right ethical path as well as anticipate future issues and potential risks that must be addressed.
- **With concerns about how data is used in AI systems being addressed by the strong data protection legal framework** instead of considering the introduction of regulation or legislation the focus should instead be on embedding an ethical approach into the development, delivery and use of AI systems and technologies.

Defining Artificial Intelligence

- 4. Before answering the questions raised by the Committee that are relevant to techUK members, it is important to define what is meant by AI and explain the different technological component of AI.**
5. Artificial intelligence “AI” is a branch of computer science that can broadly be understood as computational devices and systems made to act in an intelligent manner based on a given set of inputs. In other words, AI is a technology that can be used to enhance human decisions.
6. Different types of AI technologies are used, often together, to enable a computer to mimic the following different human behaviours:
 - **Learning** – Machine learning technologies allow a machine to learn from its experience and mistakes.¹²⁷²
 - **Reasoning and problem-solving** – Algorithms provide the instructions and steps for a machine to reach a specific outcome. For example, to beat a human in the complex Chinese game of Go.¹²⁷³

¹²⁷² <https://royalsociety.org/topics-policy/projects/machine-learning/what-is-machine-learning-infographic/>

¹²⁷³ <https://www.theguardian.com/technology/2017/may/23/alphago-google-ai-beats-ke-jie-china-go>

- **Perception and vision** - Machine vision identifies objects, patterns, faces, scenes and activities in images. Current uses include object descriptions for the blind, facial reconstructions, car-safety systems that can auto-park and detect pedestrians.
- **Language and understanding** – Natural Language processing allows computers to analyse, understand, and generate language to interface with humans. Examples of applications include transcribing notes dictated by healthcare professionals, automatically drafting text, and translating text and speech.¹²⁷⁴

The Pace of Technological change

What is the current state of artificial intelligence and what factors have contributed to this?

7. The UK has a long heritage in the development of AI technologies both in its universities and businesses.
8. The availability of high performance computing (HPC), big data and data analytics, cloud computing, internet of things and low latency connectivity is providing organisations the computing power and resources needed to develop, deliver, deploy and use AI.
9. This strong digital ecosystem underpins the UK's vibrant, competitive and thriving AI industry. A range of AI tools, technologies and solutions, increasingly offered via the cloud as a service, are available to UK organisations across both the public and private sector.
10. At the end of 2016 it was estimated that 60% of all UK AI companies were founded in the last 36 months with a new UK AI company being created on a weekly basis¹²⁷⁵.
11. The UK is becoming a global leader in establishing and growing AI companies. In a recent survey of European AI companies, 40% were based in the UK¹²⁷⁶.
12. This success is attracting investment firms looking to capitalise on the development of innovative AI technologies. In 2016 VC firm Octopus Ventures announced a £120 million fund to support UK AI start-ups¹²⁷⁷. In addition, the UK firm Dyson recently invested £2.5 billion in the creation of

¹²⁷⁴ https://en.wikipedia.org/wiki/Natural_language_processing

¹²⁷⁵ <https://medium.com/mmc-writes/artificial-intelligence-in-the-uk-landscape-and-learning-from-226-startups-70b9551f3e4c>

¹²⁷⁶ <http://tech.eu/features/13538/list-artificial-intelligence-ai-startups-europe/>

¹²⁷⁷ <http://uk.businessinsider.com/octopus-ventures-has-raised-a-120-million-fund-to-invest-in-uk-startups-2017-3>

a robotics and AI technologies centre¹²⁷⁸. This interest in the UK AI industry underlines the amount of untapped potential in the field of AI innovation.

13. Our world leading universities and research centres are at the centre of work that will shape the future of AI. The close relationships developed between universities and industry is also supporting innovative AI R&D to happen here.

How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

14. There is a wide array of predictions about how AI will develop in the future. These range from where we are today, where AI is being used as a tool to aid relatively simple processes (which some refer to as 'narrow AI'), to a future where autonomous, intelligent AI machines, or robots, are developed with humanlike mental capabilities (which is often referred to as 'general purpose or strong AI'). However, as Sir Nigel Shadbolt has stated, "We are a very, very long way away from self-aware or even generally intelligent computers"¹²⁷⁹.
15. During the course of this Parliament we will see a major shift in the growth and adoption of AI. For example, the introduction of self-driving vehicles will see autonomous AI machines that can learn, adapt to situations and make decisions without the aid of human intervention on UK roads. Looking further into the future an increase in the convergence of AI systems and autonomous intelligent machines, such as Robotic Process Automation (RPA), is also likely. Many sectors will be looking to learn from how industries such as banking, insurance, healthcare and transport adopt and use AI over the next five years. The lessons learnt from these sectors is likely to determine how the adoption of AI will develop in the longer term.
16. In terms of the impact on employment, the development and application of AI is expected to lead to the creation of additional higher-value technical roles that will be needed to service and support these advanced, autonomous AI driven machines and systems. As Jürgen Maier, Siemens CEO, has stated: "if we get this right, it doesn't just drive productivity, but it also means that you're driving jobs up the value chain"¹²⁸⁰. Experts in automation and AI with skills in areas such as software development,

¹²⁷⁸ <http://www.bbc.co.uk/news/uk-england-wiltshire-39117982>

¹²⁷⁹ <http://www.techworld.com/data/artificial-intelligence-fears-overblown-says-ai-expert-sir-nigel-shadbolt-3622100/>

¹²⁸⁰ <https://www.theguardian.com/business/2017/may/07/robotics-ai-and-3d-printing-could-close-uks-productivity-gap>

system design, engineering, programming, robotic processing automation software implementation and data science are already in high demand. In fact, all of these roles are areas where the UK currently has a domestic skills shortage.

17. The digital skills gap is one of the most urgent policy challenges facing the UK. It is estimated that the UK is already losing £2 billion a year from today's unfilled digital roles¹²⁸¹. There is real concern that a lack of skilled talent could hinder the UK's ability to realise the full opportunities offered by AI. Action must be taken urgently to create a talent pool of skilled individuals that will be needed to support the adoption and use of AI systems¹²⁸². We must also consider how we will upskill and retrain the existing workforce in years to come to ensure individuals can adapt to the new digital roles and job opportunities created by AI.
18. It is not just the digital sector where the UK may be facing a workforce shortage in the future. By 2022 it is estimated that just over a third of the current UK workforce¹²⁸³ will reach retirement age with 12.5 million jobs being vacant¹²⁸⁴. Based on the current birth rate and migration levels, studies have predicted that the UK labour pool is expected to be reduced by 2 - 3 million by 2020 and 2022¹²⁸⁵. This is leading to concerns as to whether the UK will have sufficient workers to support our growing economy and society in the future. The UK's current leadership in AI as well as automation and robotics could play a significant role in ensuring the UK's future is not hindered by a reduction in workforce numbers. If this is to be the case then it is even more important that we are developing now the talent pool of digital skills we will need to support the increased use automation of jobs using robotics and AI.

Is the current level of excitement which surrounds artificial intelligence warranted?

19. AI is creating great excitement because it will bring significant change to the way we all live, work and interact with technology. techUK believes that AI can positively benefit individuals, businesses and society as a whole. For example:

¹²⁸¹ <https://www.techuk.org/insights/reports/item/9469-the-uk-s-big-data-future-mind-the-gap>

¹²⁸² https://www.techuk.org/images/Global_Tech_Talent_Powering_Global_Britain_March_2017.pdf

¹²⁸³ <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/uklabourmarket/latest>

¹²⁸⁴ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/555606/future-of-ageing-older-workers-meeting-cowley.pdf

¹²⁸⁵ https://www.bcgperspectives.com/content/articles/management_two_speed_economy_public_sector_global_workforce_crisis/?chapter=3

- the ability to diagnose diseases like cancer quicker and more efficiently¹²⁸⁶,
- discovery of new medical treatments for diseases such as Motor Neurone Disease¹²⁸⁷
- increased business efficiency, effectiveness and productivity across sectors
- the introduction of driverless vehicles that can reduce road congestion and pollution¹²⁸⁸
- access to online products and services, including public services, that reflect the changing way we live and work
- more consistent decision making based on data not human emotion
- the identification and removal of offensive online content
- managing the constantly evolving online threat environment
- the removal of discrimination and bias in recruitment ¹²⁸⁹.

20.If the UK is to realise the full potential of these, and other economic and social opportunities offered by AI, we need both confidence and vigilance. We must be confident about the benefits AI can offer and do more to encourage organisations to embrace AI . At the same time vigilance is needed to ensure decisions made now about our AI future recognise public concerns and put the needs of humans and human values at the heart of technological innovation. techUK believe there is a role for the Committee to build consensus on what the UK’s AI driven future could look like, consider how we get there and ensure this debate is both fair and balanced.

Impact on society and public perception

How can the general public best be prepared for more widespread use of artificial intelligence? Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated? Should efforts be made to improve the public’s understanding of, and engagement with, artificial intelligence? If so, how?

21.The widespread use of AI can be hugely positive for individuals, business and society. However, the economic and social benefits offered by AI will not be fully realised if citizens do not believe this too. The goal for policy-makers and industry alike must be to ensure that the public has confidence that AI can be developed and used in a way that enhances people’s lives. We must do more to anticipate and prepare for the changes

¹²⁸⁶ <https://www.digitalhealth.net/2016/09/google-deepmind-trial-to-improve-cancer-treatment/>

¹²⁸⁷ <http://www.wired.co.uk/article/benevolent-ai-london-unicorn-pharma-startup>

¹²⁸⁸ <http://www.bbc.co.uk/news/technology-41038220>

¹²⁸⁹ <http://www.business2community.com/human-resources/4-promising-ways-ai-helping-diversity-recruitment-01907089#RbDAdX0mGRSBosix.97>

AI will bring to people's lives and build a culture of trust and confidence in AI that support its increased development and use.

22. To achieve this goal, we must move on from simply talking about the hypothetical uses for AI to demonstrating the practical reality of how AI can make people's lives easier. For example, personal mobile assistants using AI to answer real time questions and provide information; AI driven chatbots providing online customer service; AI language processing transcribing of medical notes; AI systems detecting financial fraud and to run email spam filters. AI is even being used to help fantasy football teams predict winners and losers¹²⁹⁰.
23. These examples demonstrate that AI delivers real direct benefits that we can see every day. But there are also indirect benefits that individuals and society will be benefiting from. For example
- money saved by reducing admin costs in healthcare can be redeployed to invest in more doctors and nurses providing front line care;
 - strengthening consumer protections from fraud meaning the cost of fraud is not passed onto consumers and in the near future
 - increased business productivity across sectors will lead to economic growth and job creation
 - driverless vehicles will make our cities cleaner, safer, more efficient and enjoyable places to live.

We need to do more to raise awareness and generate public interest in these and other positive benefits of AI moving forward.

24. If we are to build a culture of trust and confidence it is important that we address concerns that exist. For example, we must ally concerns that AI systems are being designed with unintentional gender or ethnic biases. A recent BBC News article explored whether a lack of women and ethnic minorities involved in designing machine learning tools may be leading to AI systems ignoring diversity¹²⁹¹. To ensure AI system reflect the society we live in more must be done now to increase diversity in those entering the computer science and AI research community. The good news is that while the number of women in IT courses remains low, AI focused courses have 28% female participation. This years A-Level results has showed a 34% increase in female students taking computing.¹²⁹² However, this is

¹²⁹⁰ <http://www.bbc.co.uk/news/technology-40905913>

¹²⁹¹ <http://www.bbc.co.uk/news/technology-39533308>

¹²⁹² <https://www.techuk.org/insights/news/item/11264-as-tech-companies-face-digital-skills-gap-a-level-results-welcomed-by-industry>

still not enough and efforts must be made to increase gender and ethnic diversity in those studying computer science, machine learning and AI.

25. We must also address concerns about the impact on jobs. According to a consumer survey 88% of people in the UK are concerned about job losses from AI¹²⁹³. There are however many differing views on the exact impact of AI on the UK labour market. Research by PwC estimates that 30% of UK jobs could be impacted by automation and AI by 2030¹²⁹⁴. It also states that the development of AI will create additional value-added roles across different industries and sectors. Similarly, a Deloitte study also predicted job losses in areas such as transport and storage, manufacturing and wholesale retail but also forecasts a growth in jobs in health and social work, education and scientific and technical roles due to automation and AI¹²⁹⁵.
26. Before we can fully address job concerns it is important that we first do the work to fully identify and understand the real economic, social, ethical and moral impact on the UK workforce of AI. It is vital that we have a realistic, constructive and balanced discussion on the opportunities and challenges AI will bring to the UK workforce. This is a task that neither Government, industry, Parliament or academia can do alone. techUK believe this is an issue that the Government's proposed Data Ethics Commission could be well placed to explore.
27. However, this debate must not focus solely how AI will replace people. Instead it should consider how the adoption of AI by organisations offers the potential to free up human resources to be used in more productive and value generating roles and accelerate the use of robotic processing automation (RPA) tools that can help humans be more efficient and effective in everyday tasks. The Committee should consider what immediate action is needed from Government and policy makers now so that we can all prepare for this future by identifying the skills needed, how to train or retrain people accordingly, and how to better connect people with opportunities.
28. As we move forward into an era where AI has an increasingly widespread and pervasive role across both the public and private sectors, trying to determine who in society will gain the most, or least, is not the right focus for this discussion. Instead the goal for policy-makers and industry alike

¹²⁹³ <http://www.webershandwick.com/uploads/news/files/AI-Ready-or-Not-report-Oct12-FINAL.pdf>

¹²⁹⁴ <https://www.pwc.co.uk/press-room/press-releases/Up-to-30-percent-of-existing-UK-jobs-could-be-impacted-by-automation-by-early-2030s-but-this-should-be-offset-by-job-gains-elsewhere-in-economy.html>

¹²⁹⁵ Deloitte (2016) *Transformers: how machines are changing every sector of the UK economy*. Retrieved from <http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-transformers-2016.pdf>

must be to ensure we find ways to bring everyone in society on the UK's AI journey.

Industry

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

29. Artificial Intelligence is expected to be worth £654 billion to the UK economy by 2035 and increase UK GDP by 10.3% by 2030¹²⁹⁶. This economic value will come from both the direct GDP growth from the tech sectors that develops, manufactures and produces AI technologies, and the indirect GDP growth by traditional sectors that adopt and use AI systems.
30. The AI market itself is expected to grow from \$8 billion in 2016 to \$47 billion¹²⁹⁷ by 2020 with Gartner predicting that every digital software product and service released by the technology industry by 2020 will include AI capabilities¹²⁹⁸. The convergence of AI with cloud computing, big data analytics and Internet of Things will drive this significant growth of the industry.
31. However, it is not just the technology sector that stands to benefit from the increased adoption and use of AI. Artificial intelligence is set to become the next big digital disrupter to the way traditional sectors operate. It is likely that the adoption of AI by companies will increase productivity, efficiencies, cost savings and overall economic growth across all industries and sectors. However, recent research by Accenture¹²⁹⁹ has found that the following key sectors are expected to gain the most from the use in AI:
 - Financial services – Using machine learning financial trading orders can be placed faster than a human trader and more frequently in multiple markets simultaneously. Banking and insurance firms are already using machine learning to detect and stop fraud.¹³⁰⁰
 - Manufacturing – AI enables the creation of smart, responsive production lines that can adapt to business changes in real time¹³⁰¹.

¹²⁹⁶ <https://www.pwc.com/us/en/press-releases/2017/report-on-global-impact-and-adoption-of-ai.html>

¹²⁹⁷ <http://www.idc.com/getdoc.jsp?containerId=prUS41878616>

¹²⁹⁸ <http://www.gartner.com/newsroom/id/3763265>

¹²⁹⁹ <https://www.forbes.com/sites/louiscolombus/2017/06/22/artificial-intelligence-will-enable-38-profit-gains-by-2035/#2db988e21969pPO>

¹³⁰⁰ <https://www.technologyreview.com/s/604122/the-financial-world-wants-to-open-ais-black-boxes/>

¹³⁰¹¹³⁰¹ <http://www.businessinsider.com/sc/artificial-intelligence-change-manufacturing?IR=T>

- Retail – Natural language processing is already being used to enable customer service chatbots that can respond to questions and requests anytime of the day¹³⁰².
- Transport – AI is underpinning the development of autonomous vehicles including cars, planes, lorries and ships that will reduce energy consumption, reduce congestion and increase the efficiency of delivery services.
- Professional services – The use of AI in the legal sector can today turn vast amounts of data into insights and knowledge that can be used in court cases¹³⁰³.
- Healthcare - Connected AI enabled medical devices will analyse information in real time and make decisions as to when human intervention is needed. The pharmaceutical industry is using high performance computing combined with AI to discover new drugs¹³⁰⁴.
- Construction – Using machine vision for surveying building sites and autonomous machines to support humans move heavy machinery safely and driving efficiency in tasks such as bricklaying.¹³⁰⁵

32. In the labour intensive services sector, such as retail and construction, the use of AI will see the removal of highly repetitive administrative or manual tasks, leaving staff to interact more with customers. This will help companies to build brand loyalty and increase competitiveness and revenues at a time when the UK services sector has been showing signs of slowing down.¹³⁰⁶

33. For sectors where capital expenditure is high, such as manufacturing and transport, the use of AI could see significant increase in efficiency, cost savings and profits by being able to predict and prevent the failures of equipment.

34. While there are business benefits for organisations from using AI, a recent business survey indicates that there is still more that needs to be done to encourage organisations investment and adoption in AI. According to the CBI 47% of UK firms currently have no plans to invest in AI in the future¹³⁰⁷. It is hoped that the Government's independent AI Review currently being conducted will offer recommendations on how to support

¹³⁰² <http://www.insider-trends.com/what-chatbots-mean-for-retail/>

¹³⁰³ https://en.wikipedia.org/wiki/Electronic_discovery

¹³⁰⁴ <https://uk.reuters.com/article/us-pharmaceuticals-ai-gsk/big-pharma-turns-to-ai-to-speed-drug-discovery-gsk-signs-deal-idUKKBN19N003>

¹³⁰⁵ <http://www.conexpoconagg.com/news/october-2016/ai-and-robotics-the-future-of-construction/>

¹³⁰⁶ <https://www.theguardian.com/business/2017/jul/05/uk-services-sector-growth-hits-four-month-low-amid-brexit-fears>

¹³⁰⁷ <http://www.cbi.org.uk/news/half-of-firms-expect-ai-to-transform-their-industry/>

and encourage UK organisations across all sectors to invest in and adopt AI.

35.techUK also sees huge opportunities in the use of AI to support the delivery of increasingly digitised, personalised and responsive public services. Recently Enfield Council's has deployed an autonomous virtual agent called Amelia¹³⁰⁸ that analyses and understands natural language and context to apply logic and learning to resolve citizen problems. However, this example of adoption seems rare at the moment even though it is estimated that the use of AI virtual agents across Government departments and the public sector could save an estimated £4 billion a year¹³⁰⁹.

How can the data-based monopolies of some large corporations, and the 'winner takes-all' economies associated with them, be addressed?

36.This should be a matter for competition law. Competition authorities should ensure that they have the knowledge and resources necessary to understand the data driven economy.

11. How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

37.In May 2018 the new EU General Data Protection Regulation (GDPR) will enter into force. This will ensure there is a strong legal framework in place, including significant sanctions, to ensure individuals data is protected, managed and secured in light of the use of technologies such as machine learning and AI.

38.As not all data will be personal data, it is also important that organisations have in place appropriate measures to ensure the integrity, confidentiality, privacy and security of non-personal data. To achieve this, it is suggested that organisations follow three key steps:

1. Put in place data governance policies and procedures based on relevant legal and regulatory requirements. For personal data this would be the requirements of the new EU General Data Protection Regulation that enters into force in May 2018.
2. Adopt appropriate technological tools and solutions that can protect the integrity, validity and security of data throughout its lifecycle.
3. Ensure employees are given training and have the right skills to manage and keep data secure.

¹³⁰⁸ <http://www.ipsoft.com/2016/07/18/first-public-sector-role-for-amelia-as-enfield-council-deploys-her-to-boost-local-services/>

¹³⁰⁹ <https://www.theguardian.com/technology/2017/feb/06/robots-could-replace-250000-uk-public-sector-workers>

39. As data continues to play a bigger role in public and private sector bodies decision making, ensuring the validity of the data will become increasingly important. Big data analytics and machine learning technologies work by bringing together vast amounts of structured and unstructured data that is analysed to find hidden insights, knowledge or the answer to a specific question. The results of this analysis can then be used to make decisions that might impact the economy, society or an individual's life. The quality of the data being used to make this decision may therefore determine the quality of the decision reached; good data in means good data out. The ability to demonstrate the validity, quality and integrity of data, including a lack of bias within data sets, could become increasingly important in determining whether the decisions made by autonomous, algorithmic driven AI systems can be trusted by organisations and the general public.

Ethics

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

40. We are entering a new era where AI is at the very heart of future technological innovation. This is raising profound social and ethical questions about how data is used that go beyond the legal framework for the protection of personal data. For example:

- What does it mean to be human and the distinction between humans and machines?
- To what extent we will be content as a society to transfer responsibility and control in certain situations from a human to a machine?
- How do we ensure AI systems are doing what they are supposed to be doing?
- How are we able to verify AI systems safety and ensure they do not malfunction or are vulnerable to cyber attacks?
- Are decisions made by AI auditable, challengeable and ultimately understandable by humans?

41. As the pace of AI innovation continues to accelerate techUK believe the time is right to put mechanisms in place that bring together academia, business community, Parliamentarians, policy makers and others to discuss these and other emerging issues in the future. We need a way to ensure technological innovation is on the right path, anticipate future ethical implications raised by the development of AI and mitigate any potential future risks.

42. techUK welcomed the recommendation in the recently published Royal Society and British Academy report on Data Management and Use:

Governance in the 21st Century¹³¹⁰ calling for the creation of a new data stewardship body¹³¹¹. This body could be a significant step forward in building the capability and capacity we will need to anticipate future ethical issues and put in place effective safeguards that ensures AI systems are developed based on human values and act in the interests of humans. techUK has also supported the recent announcement by the Nuffield Foundation to create an independent Data Ethics Convention in 2018 and looks forward to supporting this initiative¹³¹².

43.techUK has also urged the Government to follow through on its 2017 General Election manifesto pledge to establish an independent Data Use and Ethics Commission to “advise regulators and parliament” on data use issues¹³¹³. Any Commission created must have a broad membership of people who think about ethical data issues in different ways and must be given a remit that is long term and extends beyond a single Parliamentary session if the Commission is to advise Parliament on future developments in AI.

44.The initiatives in this area are supported as it offers an opportunity to position the UK as a global leader in identifying, understanding and discussing ethical issues being raised by the development of AI. This would support the Government’s post Brexit Global Britain ambitions by encouraging global companies looking to develop and use AI to come to the UK to seek advice and support on how to address ethical issues.

In what situations is a relative lack of transparency in artificial intelligence systems (socalled ‘black boxing’) acceptable? When should it not be permissible?

45.Transparency, openness and accountability all have an important role to play in building trust and confidence in the use of AI. However, the current public debate on the transparency of AI systems often focuses more on the need for organisations to open up AI systems and show the algorithm, or computer code being used.

46.The release of hundreds of lines of highly complex computer software code will not increase transparency or understanding about an AI system. This code is unlikely to be understood by the general public and, for some systems, even those with the appropriate programming language knowledge. Instead the focus of this discussion should be on how to

¹³¹⁰ <http://www.techuk.org/insights/news/item/10993-why-data-governance-must-keep-pace-to-secure-uk-s-ai-future>

¹³¹¹ <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf>

¹³¹² <http://www.nuffieldfoundation.org/news/nuffield-foundation-announces-additional-%C2%A320-million-research-funding-fellowship-programme-and->

¹³¹³ <https://s3.eu-west-2.amazonaws.com/manifesto2017/Manifesto2017.pdf>

ensure that the right mechanisms are in place so that the decisions and outcomes made by AI systems are transparent, fully understandable and open to challenge and redress by both businesses and consumers.

47. It should also be remembered that for companies developing AI systems the algorithms used are a crucial part of businesses intellectual property and will be commercially confidential and protected by Trade Secret Law¹³¹⁴. How AI systems are applied and used in specific sectors, particularly in defence and areas of national security, will also be subject to strict contractual non-disclosure and confidentiality agreements.
48. There is concern that any recommendations in this area could result in AI companies avoiding the UK due to future competition concerns. This would place UK businesses wanting to benefit from innovations in AI at a competitive disadvantage and endanger the Government's ambition to encourage AI companies to thrive and grow in the UK.

The role of the Government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom?

49. Government has a key role to play in ensuring a supportive legal, regulatory and fiscal environment that enables UK AI companies to develop and scale, encourages public and private sector AI use and attracts global AI companies to invest in the UK in a post Brexit world. Government can also facilitate further dialogues with industry, academia, civil society, and other interested stakeholders to help shape the development of AI innovation to achieve its potential.
50. The recent financial investment in AI outlined in the Government's Industrial and Digital Strategies have been welcomed by industry. The £93 billion AI Industrial Challenge Fund¹³¹⁵ and the £17.3 million to support UK universities research means the UK can continue to define the future by being a European and global leader in AI R&D¹³¹⁶.
51. While this investment is important, the Government should also look to support digital entrepreneurs looking to develop new AI business models. The introduction of R&D tax credits for SME's developing AI could encourage AI entrepreneurialism and support further development of the UK's AI industry.
52. The Government could also play a leadership role in encouraging AI adoption across all sectors including the public sector. For example, it is

¹³¹⁴ <http://www.robinskaplan.com/resources/articles/software-and-trade-secrets-rethinking-ip-strategies-after-cla-v-alice>

¹³¹⁵ <https://www.gov.uk/government/news/business-secretary-announces-industrial-strategy-challenge-fund-investments>

¹³¹⁶ <https://www.gov.uk/government/news/17-million-boost-for-the-uks-booming-artificial-intelligence-sector>

suggested that every Government department should consider, and then report, on the benefits of using AI technologies to support civil servants in meeting public service demands. In particular how and where AI could help to deliver the Government's Digital Transformation Strategy goals by 2020.

53. Data is fundamental to the development and use of AI systems. The continued free flow of data between the UK, Europe and the rest of the world is therefore vital to the UK's AI future. As the UK prepares to leave the EU the Government must ensure that a mutual data adequacy agreement with the European Commission, is in place before we leave so data can continue to flow between the UK and the EU.

Should artificial intelligence be regulated? If so, how?

54. Today many of the concerns regarding the use of AI technologies, such as algorithmic driven machine learning, are focused around how data is being used in these systems. It is important to remember that the current data protection legal framework is already sufficient to address concerns around how personal data is being used. In fact, this legal framework has recently been updated and strengthened to anticipate the increased use of machine learning and AI.
55. The new General Data Protection Regulation (GDPR) will come into force in the UK on 25 May 2018. techUK sees GDPR as a significant step forward in ensuring there is a strong legal framework in place to ensure individuals data is protected in light of technologies such as AI. The GDPR increases individual's legal rights and choices where solely automated decision processing that has a legal effect takes place. As the ICO has stated, the GDPR introduces "stricter rules" for personal data that are "no different for big data, AI and machine learning"¹³¹⁷.
56. There are other concerns about AI that are perhaps less focused on the use of personal data. These include algorithmic transparency, the safety of autonomous systems, and the broader societal impact of AI. A recent Royal Society report on Machine Learning in fact explored the role of algorithms and identified how the GDPR could also help to address transparency concerns¹³¹⁸. Where there are other concerns about how AI is developing these need to be fully identified, understood and discussed before determining whether regulation or legislation has a role to play. As the House of Commons Science and Technology Select Committee's own

¹³¹⁷ <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>

¹³¹⁸ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

inquiry report recently stated, “it is too soon to set down sector-wide regulations for this nascent field”¹³¹⁹.

57. At this stage of AI’s development techUK believes the right approach to take is one that focuses on embedding an ethical approach into the development, delivery and adoption of AI systems and technologies. The Government should also focus its efforts on how to enable broad deployment of AI, and its continued innovation in every sector, including the public sector.

7 September 2017

¹³¹⁹ <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>

Thames Valley Police – Written evidence (AIC0125)

Select Committee Submission on the Implications of Artificial Intelligence (AI)

Submitted by the Strategic Governance Unit on behalf of Thames Valley Police

Contents

- Definition 1441
- Pace of Change 1441
 - The current state of AI? 1441
 - Is the current level of excitement warranted? 1441
- Impact on Society 1441
 - How can the public be best prepared?..... 1441
 - Who is gaining the most/least? 1442
- Industry (Organisational) 1442
 - What sectors stand to benefit?..... 1442
- An Idealistic Imagining of the Future of AI Policing 1443
- Ethics 1443
 - What are the ethical implications? 1443
 - In what situations is black-boxing acceptable? 1443
- The Role of the Government 1444
 - Should AI be regulated?..... 1444

Definition

This submission defines Artificial Intelligence as machine functions designed to mimic human cognition. Machine Learning is defined as a process where the machine develops its own decision making model based upon the data it is exposed to, and this is achieved using “black box” algorithms whereby only the input and output can be viewed and analysed. This submission considers AI Types I & II only, in reference to policing.

Pace of Change

The current state of AI?

AI is only starting to be used in UK policing. Examples include the Durham Custody Harm Assessment Risk Tool (HART) and facial recognition used by the Met at Notting Hill Carnival. There are multiple levels to AI capability and even at the lowest level, AI could perform many of the process driven tasks that take place in the police. The scope for AI in policing is huge. Police forces are introducing contact management systems that allow the public to have a greater online relationship with the police. This includes the reporting of crimes and the submission of electronic evidence. Early indications are that the police will face a deluge of electronic material. Intelligent content management will be one of the most impactful implementations of AI in policing.

Is the current level of excitement warranted?

The current level of excitement around AI in policing is warranted. As more complex digital data sources become available, it will become prohibitively complex to conduct investigations. AI will be necessary for automating functions that are too time consuming or complicated for humans to do. Examples of this include assisting investigations by ‘joining the dots’ in police databases, the risk assessment of offenders, forensic analysis of devices, transcribing and analysis of CCTV and surveillance, security checks and the automation of many administrative tasks. This will fundamentally change police functions and the resourcing of those functions. It may be necessary for the Government to lead on the design of a national policing AI framework, as the data integrity and lack of interoperability in disparate police force ICT infrastructures may prohibit the implementation of effective AI systems.

Impact on Society

How can the public be best prepared?

There is an expectation from the public that the police will do everything in their power to keep citizens safe and disrupt crime and this has to be balanced with expectations of privacy. Transparency will be vital in ensuring legitimacy, and there needs to be an open discussion regarding the techniques used by the police, except in situations where that information will compromise capability.

There is also the increased likelihood of criminals exploiting AI for their own purposes. This may lead to types of criminality that are difficult to foresee.

Who is gaining the most/least?

In reference to policing, those who are likely to benefit from AI are the public. The ability to quickly identify offenders, victims, locations of crime, and trends and patterns will result in improved police performance at decreased cost. The police service would need to undergo significant change, and depending on the requirement to make productivity savings, funds would need to be redeployed to priority areas. It is essential that the workforce is willing to adapt to change and learn new skills so that they can be retained in the roles that AI cannot perform.

Industry (Organisational)

What sectors stand to benefit?

Below is a short list of the most prominent areas where AI could improve policing.

Crime Reporting / Recording Processes

- Demand Monitoring
- Intelligent Incident Categorisation
- Automated allocation to the appropriate emergency service
- Automated Incident / Crime Recording

Crime Prevention & Investigation / Multisource Data Analysis

- Content Management
- Suspect / Victim Identification
- Threat, Risk and Harm Assessment (Prediction)
- Predictive / Adaptive Patrols
- Crime Solvability Factors
- Open-source Analysis (e.g. social media hate crime)
- IoT Integration

Video / Image / Audio Analysis

- Facial / Emotional / Gait / Behavioural / Crowd Flow Recognition for all forms of video such as Body Worn Video, CCTV evidence and video submitted by the public.
- Verbal Statements / Interviews & Questioning / Speech analysis

Back Office Functions

- Check Payment
- Invoicing
- Police Officer Applications / Vetting Processes
- Facilities and ICT Services
- Licence Applications

An Idealistic Imagining of the Future of AI Policing

A member of the public calls 999 and describes an ongoing incident to an AI system. Speech analysis categorises the type of incident and detects indicators of stress from the caller. The date, time, location and offence details are recorded automatically onto police systems. Connected CCTV and IoT systems would already be monitoring the disturbance, and by using behavioural recognition and face recognition have identified the suspect and the victim who have both previously come to police attention.

Risk assessments are automatically conducted on both parties using police and partner data. The offence is prioritised against other ongoing incidents and the system notifies a unit with the appropriate skills to deal with the situation, which happens to already be nearby on a predictive patrol pattern.

The suspect is arrested and all parties, including the arresting officers provide voice statements in situ, which are automatically uploaded, transcribed and attached to the crime report. Solvability factors are calculated on the quality of the available data. The risk assessment provides a recommendation for officers on the next steps for the offender and also an appropriate support package for the victim.

Ethics

What are the ethical implications?

We need to instil an evidence-based approach to the applications and outcomes of AI. Collaboration with academia will be necessary to ensure that the development of AI technologies result in fair outputs. Recent tests of AI in policing indicate there is a risk of bias perpetuation in AI outputs, therefore engagement with Privacy and Civil Rights groups will be necessary to persuade the public that everything possible is being done to mitigate this whilst doing our best to keep them safe. Of utmost importance is that any AI process that involves an ethical issue, must have a high level of human oversight and clear justification. The automation of processes also introduces a risk of being unable to reason with a human when events occur outside expected parameters. Ethics committees will undoubtedly be required to rigorously chart AI processes and frameworks will need to be in place to properly deal with events of a non-standard nature.

In what situations is black-boxing acceptable?

By definition, machine learning algorithms are a black box. In much the same way as a calculator uses algorithms to provide an answer given an input of a calculation, machine learning provides an output that a human technically may be able to generate, but at far greater speed and efficiency. The other benefit of machine learning, which is not always possible in other methods or even with human decisions, is that the error rate and performance of the model is provided, giving a degree of confidence to the results.

Rather than being deemed acceptable or unacceptable, black boxing is currently unavoidable. Efforts need to be made to ensure that if an AI is in a position to cause harm, or is involved in any judicial decision making process, that it is technologically possible to derive an explanation for how the output was formulated. Where this is too complex, it may be that AI can only be used to support decision making. Machine learning algorithms are a useful tool but we need a full understanding of how the methods work and thorough policies on how to use them appropriately.

The Role of the Government

Should AI be regulated?

To ensure that AI development progresses in a responsible manner, the use of AI should be regulated using the Asilomar AI principles as a starting point. This is important when it comes to safety, security and judicial transparency.

It is also likely that criminals and unscrupulous businesses will find ways of exploiting AI for their own gain. Legislation will have to be drafted to ensure there are sufficient penalties for the improper development and misuse of AI.

6 September 2017

Thomson Reuters – Written evidence (AIC0223)

Please permit me to submit two papers (attached) on the topics of ethics in the broad field of artificial intelligence (more specifically: Natural Language Processing, NLP) for your consideration.

They were published at the First Workshop on Ethics and NLP in Valencia, Spain, April this year, and suggest *Ethics Review Boards* and *Ethics Check-Lists* be used to document and deal with ethical issues arising in the context of research in Artificial Intelligence-related disciplines and their commercialization.

The emerging fields of machine learning, natural language processing and other sub-fields of AI are advancing at a rapid pace, and the ethics and regulatory debate surrounding these technologies has barely started.

Companies are commercially incentivized to apply AI broadly, and I expect it to be beneficial, indeed necessary, to set regulatory boundaries and to install oversight instruments to avoid the following:

- Unethical AI Applications: Uncontrolled abuse of data to unethical ends; an example is the automatic machine detection of sexual orientation from photographs/CCTV recently conducted at Stanford University, which deprives individuals of their right to privacy, and to keep their most intimate properties to themselves if they so wish;
- Systematic discrimination caused by improper data selection where data is used in systems that make use of machine learning; machine learning-based systems require training, which is often done by data from white males only, for unbiased data collection would be more costly and take much more time. However, machine learning methods only work well for cases represented in the training data. Non-white, non-male, or other excluded minority groups would not be processed properly by machine learning systems;
- Minorities could be unfairly disadvantaged by being excluded from access to essential services. Imagine a voice recognition system to do your banking over the phone, as banks are reducing physical branches. Such a system would likely be trained with British voices available in London if the company developing the system is London-based. Said system likely will result in misrecognitions, or may not work at all, for an elderly citizen in Uddingston, Scotland, and lacking alternatives access to cash will depend on trusted friends or family members, if available. For economic reasons, in the absence of regulations, no company will undertake training a speech recognizer with all diverse groups residing in the UK (in the related case of accessibility of Web pages, regulations have been a success).

There are elements lobbying for a 'free reign of technology', arguing it is 'too late' to stop technological 'progress'; however, I disagree with such a *fatalistic*

attitude: all science must be conducted, and technology be exploited, in the service of humanity. Laws and regulatory controls can prevent or limit, prevent and roll back unethical use.

Should you have any questions with regards to the points made above, the papers attached, or require assistance in the assessment of regulatory models for practical feasibility from an industry perspective, I would be most happy to answer any questions you may have.

Dr Jochen L Leidner, *MA MPhil PhD*

(Writing in a capacity as a scientist and a Director of Research,
Research & Development, Thomson Reuters)

Attached:

- 2 scientific publications on ethics in the realm of natural language processing

Please note that these publications are not included as evidence, but they have been circulated to the Committee Members.

22 September 2017

Touch Surgery – Written evidence (AIC0070)

DIGITAL SURGERY

House of Lords AI Enquiry Submission, Touch Surgery

SEPTEMBER 2017 / AUTHORS FROM TOUCH SURGERY

We believe the key opportunities for AI in surgery are to capture the models and trends that data can reveal that are not obvious to us because we are not exposed to the temporal or spatial information cues.

OPPORTUNITIES FOR AI IN SURGERY

AI technologies can help to assist us in discovering best practice. For example by establishing links between data, which is now commonly available (through video, patient and instrument sensors, etc.), and higher level abstractions about risk, complications and eventually by linking such data procedural outcomes. Conventional approaches to modelling the relationship between data and outcomes or risk have been limited by modelling assumptions and lacked robustness. AI can potentially overcome this limitation.

Optimising processes to achieve optimal hospital operation both on the micro and macro scales. AI can help to inform scheduling systems, updating them with real-time sensor feeds, to better utilise resources and availability. This can be implemented on the overall health service (e.g. across hospitals' operation) or on local theatre use and ward occupancy (device use, bed availability, etc.).

Evidence based healthcare, leveraging big data rather than individual trials' data. Scaling data models to large, population level studies can provide deeper insight into best practice for surgical outcomes and for pharmaceutical alternatives as well as treatment pathways. There is a significant opportunity for the UK here with the availability of NHS data but it must be capitalised on carefully to ensure long term sustainability.

RISKS FOR AI IN SURGERY

Data, privacy and protection of human and societal rights. Ultimately current AI technologies are at the root, driven by loss functions that are human driven. These may be still subject to the difficulties of modelling difficult questions about optimality. In essence the AI system will perform a search for an optimal solution better than we can, however, the criteria is currently deterministic and set by us. Lack of social or other considerations given that cost models at the root of AI algorithms are still model driven. Optimisation function needs to be carefully designed and selected.

The availability of data is a significant resource for AI driven exploitation by commercial organisations. National resources, for example handled through the NHS or other systems, should be carefully managed to encourage national benefit.

Model transferability. The ability for AI algorithms to be transitioned from one scenario or situation to another is still a major challenge. For example systems trained in one geographic or demographic location, may not apply generally to another, similarly for procedural processes. Therefore significant effort is still needed in adaptability, domain transfer and the ability to train systems from simulation or semi-real data.

4 September 2017

Transport Systems Catapult – Written evidence (AIC0158)

Contribution from: Transport Systems Catapult <https://ts.catapult.org.uk/>

Contribution authors: Zeyn Saigol and Ecaterina McCormick

Date: 6th September 2017

0. About the Transport Systems Catapult

- 0.1. The Transport Systems Catapult is one of 10 'not-for-profit' research and innovation centres in the UK.
- 0.2. We work with companies, universities and government using technology and novel commercial models to transform transport and we call this 'Intelligent Mobility'.
- 0.3. We are working to identify future technology trends so we can make the UK a global leader and create jobs and grow companies.
- 0.4. Through identifying new technology and ways of working, we are making transport better for everyone.

1. What is the current state of artificial intelligence? How is it likely to develop over the next 5, 10 and 20 years?

- 1.1. AI was born over 60 years ago at the 1956 Dartmouth workshop held at Dartmouth College, NH, USA. Since then there has been steady technical progress, but AI has fallen in and out of fashion, with corresponding variability in overall funding, several times.
- 1.2. Originally it was believed it would be easy to create a machine with the same level of intelligence as a human. This has proven far harder than expected, and that goal is often now referred to as creating an AGI – artificial general intelligence. What AI has achieved is human or better-than-human performance in a large range of limited domains, including some domains (such as chess and, more recently, go) that were expected to be almost impossible for machines.
- 1.3. Today AI is composed of many subfields, the most prominent of which is machine learning (ML). Machine learning means extracting regularities from data to form a model; for example, given many examples of handwritten digits, each annotated with the number it represents, create a model that can classify which digit a novel sample of handwriting corresponds to.
- 1.4. Within ML, deep learning (also known as deep neural networks, often implemented as convolutional neural networks) has shown itself to be highly effective on several benchmark problems. Deep learning has come to prominence recently due to the ability to store and query huge datasets, advances in computing power, and improvements in the

learning algorithms. The trade-off is that, compared to other ML methods, deep learning requires a larger amount of training data.

- 1.5. Other AI subfields include automated planning, manipulation and grasping, natural language processing, computer vision, evolutionary algorithms, knowledge representation, reasoning, and multiagent systems. There is significant crossover with other fields, such as robotics, control, operations research, and mathematical optimisation.
- 1.6. AI is now used in many “everyday” technologies, including web search, digital assistants in mobile devices, speech recognition systems used in call centres, and targeted online advertising.
- 1.7. Significant future progress in AI will be made by (a) applying ML to the problems in other AI subfields, and (b) combining ML with methods from other AI subfields. An example of (a) is using ML for object recognition in computer vision: this has enabled better-than-human performance on some tasks¹³²⁰. An example of (b) is deep reinforcement learning, which combines deep learning and reinforcement learning to learn how to act in the world, as opposed to simply how to classify data.
- 1.8. It is likely that expert knowledge of ML will become less important to successfully applying it. This is because better tools will become available to automate the analysis of a particular domain and dataset, and set up the ML models appropriately.
- 1.9. It is hard to predict timescales for advances in AI. However, it seems almost certain that AI will be used in an increasing number of practical applications in the future.
- 1.10. Opinions vary on the impact of AI on jobs. Many commentators have pointed out the potential for AI to replace up to half of the tasks currently performed by human workers¹³²¹. Others argue that a similar decline in employment was predicted prior to the industrial revolution, and the IT revolution more recently; but these turned out to produce changes in the type of work people do, rather than the total employment level.

2. Is the current level of excitement surrounding artificial intelligence warranted?

- 2.1. Artificial Intelligence investment has turned into a race between major companies.
- 2.2. Tech giants including Baidu and Google spent between \$20B to \$30B on AI in 2016, with 90% of this spent on R&D and deployment, and

¹³²⁰ ImageNet (<http://image-net.org>) object identification results, <https://arxiv.org/pdf/1502.01852.pdf>

¹³²¹ http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

10% on AI acquisitions. U.S.-based companies absorbed 66% of all AI investments in 2016. China was second with 17% and growing fast¹³²².

- 2.3. There is evidence to suggest AI's current level of interest isn't just temporary; for example, investment in AI tripled from 2013 to 2016.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

- 3.1. Education (see paragraph 5).
- 3.2. Job re-training (see paragraph 10).
- 3.3. Popular science programmes aimed at demystifying AI models. The media disseminates only the applications of AI while making it difficult for the public to understand its internal workings.

4. Who in society is gaining the most from the development and use of artificial intelligence? Who is gaining the least?

- 4.1. The segment of population negatively affected by the advent of AI is likely to be less affluent with lower levels of access to technology. While more systems are automated and lack human interaction, some people might struggle to use the new interfaces.
- 4.2. The corporations holding large amounts of data are most likely to benefit from the AI development. Such entities have a clear advantage in training their models, getting better insights and progressing faster than anybody else.

5. Should the public's understanding of, and engagement with, artificial intelligence be improved?

- 5.1. The mysticism surrounding AI is a barrier in allowing people to understand how it can affect their lives. There is a clear media focus on the negative effects and not enough emphasis on the positive effects.
- 5.2. There is a need to educate the public about the basic concepts of AI – for example, a 2015 study by the Transport Systems Catapult indicated that only 39% of people would be prepared to use an autonomous car¹³²³. Being better informed should help allay some of their fears about the adoption of the technology.

6. What are the key industry sectors that stand to benefit from the development and use of artificial intelligence?

¹³²² "Artificial Intelligence, The Next Digital Frontier", a McKinsey Global Institute Study, and discussion paper,

[http://www.mckinsey.com/~media/McKinsey/Industries/Advanced Electronics/Our Insights/How artificial intelligence can deliver real value to companies/MGI-Artificial-Intelligence-Discussion-paper.ashx](http://www.mckinsey.com/~media/McKinsey/Industries/Advanced_Electronics/Our_Insights/How_artificial_intelligence_can_deliver_real_value_to_companies/MGI-Artificial-Intelligence-Discussion-paper.ashx)

¹³²³ "Traveller Needs and UK Capability Study", Transport Systems Catapult.

<https://ts.catapult.org.uk/wp-content/uploads/2016/04/Traveller-Needs-Study-1.pdf>

- 6.1. High tech, communications, financial services, transportation and logistics, media and entertainment, health services, automotive and assembly, tourism and retail are sectors leading in AI adoption, as identified in the McKinsey report referenced previously.
- 6.2. The transport sector will benefit from AI in several key areas:
 - 6.2.1. Autonomous vehicles, both road-based and in other environments. AI has been central to making autonomous cars possible, as it enables the world to be understood through sensor data. AI will probably also be important for decision making in autonomous vehicles.
 - 6.2.2. Route planning, as already used in sat-nav systems, is based on AI methods and makes life easier for many motorists. In the future, route planning is likely to take more account of the intentions of other vehicles using the road network. Advances in similar AI algorithms will also be useful for fleet operators and mobility-as-a-service.
 - 6.2.3. There is potential for efficiencies across the sector by applying ML to the massive quantity of data generated by the industry.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economics associated with them, be addressed?

- 7.1. Anti-monopoly laws, incentives to open data sets such as those recommended in the recent Transport Systems Catapult report in partnership with the Open Data Institute, and Deloitte.¹³²⁴, a stronger collaboration between the private and public sector (Uber recently made public its aggregated data), and continuous release of open source libraries are suggestions on how to stop entities gain an unfair advantage over others.

8. What are the ethical implications of the development and use of artificial intelligence?

- 8.1. A potential danger is incorporating human biases into the AI models.
- 8.2. The usage of inaccurate data sets would introduce noise in the models and eventually generate erroneous results.
- 8.3. Due to AI models' nature, there is an emphasis on correlation and less on causation.
- 8.4. Random mutations are a strong driver for organisms' evolution. Building a system led by strict statistical models may go against natural processes and negatively impact how the society evolves.

¹³²⁴ "The case for Government involvement to incentivise data sharing in the UK Intelligent Mobility sector", Transport Systems Catapult. <https://s3-eu-west-1.amazonaws.com/media.ts.catapult/wp-content/uploads/2017/04/12092544/15460-TSC-Q1-Report-Documents-Suite-single-pages.pdf>

- 8.5. In some cases, AI may enable machines to take decisions previously only made by humans. An oft-cited example is the *trolley problem* for autonomous cars (although we believe that AI technology is a long way from being sophisticated enough to make such decisions).
- 8.6. The adoption of new technologies often involves a trade-off – for example, heavy construction machinery allows far greater productivity, but can result in accidents that would not have happened with traditional hand tools. In this respect AI is no different (although in the case of autonomous cars, it is expected they will be far safer than human-driven ones overall).

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable?

- 9.1. With most ML technologies, it is hard to understand why a sample is given its classification.
- 9.2. This may not be ideal in certain domains; for example, reviewing why a job applicant was rejected in an early-stage applicant filtering system, or understanding why an autonomous vehicle chose a path that led to an accident.
- 9.3. This is not very different from non-AI software: the onus is on the creators of the software to design it responsibly. In the case of ML software, the creators should choose the training data and input features carefully, and these should provide sufficient explanation for decisions made by the model.
- 9.4. Overall, at least for non-military domains, we feel any lack of transparency is an implementation issue, and should not be a barrier to the adoption of AI.

10. What role should the Government take in the development and use of artificial intelligence in the UK?

- 10.1. Given the potential economic benefits, the Government should promote the development and use of AI.
- 10.2. Education is key to this. Education about AI should start in secondary school, in the same way that children should gain a basic grasp of physics. This will also lay the foundations for studying the subject in further education.
- 10.3. The Transport Systems Catapult has produced a report on the skill shortage in Intelligent Mobility¹³²⁵.
- 10.4. Investment in universities is critical if the UK is to keep pace with the rest of the developed world in AI skills. Careers in academia in the UK

¹³²⁵ “Intelligent Mobility Skills Strategy”, Transport Systems Catapult. https://s3-eu-west-1.amazonaws.com/media.ts.catapult/wp-content/uploads/2016/10/31095944/3383_IM-Skills_Business-Case_Brochure.pdf

should be made attractive compared to working as an academic in another country. Funding for AI research should be boosted.

- 10.5. The Government should incentivise data sharing, as detailed in a Transport Systems Catapult report, previously referenced in 7.1. There is a delicate balance to be struck between allowing useful data to be shared, and protecting the privacy of individuals. If other countries are less concerned about privacy in data sharing, it could enable them to make faster progress in applying AI.
- 10.6. As noted in paragraph 5, educating the public is important.
- 10.7. AI will probably lead to changes in the jobs people do. The Government can, firstly, help set expectations that changes in jobs and careers are to be expected, and not something to worry about. Secondly, the Government should provide assistance for adults (including white-collar workers) to acquire significant new job-related skills during their working life.

11. Should artificial intelligence be regulated?

- 11.1. On a practical level, regulating AI would be difficult because whether or not a particular piece of software counts as AI is a matter of opinion. Early work done by the AI community has often been adopted into “mainstream” computer science as it has matured¹³²⁶. Even ML does not have clear boundaries.
- 11.2. Certain domains where AI is used, or might be used in the future, should probably be more regulated. For example, the Government may wish to introduce a test process to help assure that the software controlling an autonomous car is safe.
- 11.3. There are privacy issues relating to the storage of data, and ethical issues relating to potential biases in decision-making software. If appropriate, regulations could be formulated to address these regardless of whether AI was used in the software.
- 11.4. Some commentators believe ongoing progress in AI will eventually lead to an AGI, smarter than humans, and this could be catastrophic for the human race¹³²⁷. We believe this possibility is quite remote, and therefore regulation to guard against it would not be warranted at this point.

12. What lessons can be learnt from other countries or international organisations in their policy approach to artificial intelligence?

- 12.1. In 1945, adult literacy rate in South Korea was just 22% while today most children graduate from high school, and of these, 82% go on to university. Presently, South Korea is the fourth largest economy in Asia

¹³²⁶ For example, graph search algorithms such as Dijkstra's algorithm and A* are a well-established part of computer science, but can also be viewed as AI algorithms.

¹³²⁷ Elon Musk and others have invested in Open AI (<https://openai.com/>), which conducts research into how to create a “safe” AGI.

and the eleventh in the world. In preparing its citizens for the future of AI, it heavily invested in strengthening its whole school age education system. The incubation of the talent pool starts from primary school.

- 12.2. The creation of alternative education systems of the likes of Ecole 42 in France and the US are adapted to the current market requirements. The school doesn't have professors, doesn't issue diplomas or degrees, and is open 24/7. The learning process fits the corporate models of peer to peer pedagogy and project based learning. This system encourages students to carry out their own research and learn based on trial and error.
- 12.3. The fast development of the Chinese Silicon Valley is a clear example on how the Chinese government is supporting new kinds of applications and new waves in innovation.
- 12.4. The strong collaboration between US academia and the private sector shows how AI development can receive a meaningful boost.

6 September 2017

Richard Tromans – Written evidence (AIC0227)

Submission of Responses to SELECT COMMITTEE ON ARTIFICIAL INTELLIGENCE Call for evidence.

KEY FOCUS: AI in the Legal Sector, Wider Government Need to Invest in Training and Education as AI takes Low Skilled Jobs.

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

The current state of AI, or narrow AI, in the legal field has had a slow start, but in the last 18 months has seen exponential growth.

Research by Artificial Lawyer [*the leading publication on AI and the law*] that it has carried out in partnership with The Times Law section, shows that most of the Top 30 law firms in the UK by revenue are using AI for client work, or piloting it with a view to doing so.

This is a major change on the position just as recently as 2015, where the number would have been tiny.

Also, interest among smaller law firms in AI has also grown considerably, with firms handling consumer level matters showing strong interest in exploiting the technology. Although, price remains a barrier for smaller firms.

2. Is the current level of excitement which surrounds artificial intelligence warranted?

In the legal field, yes. What we are seeing is what I call the 'industrialisation of cognition', which is seen in the legal world by a 'New Wave' of legal tech. New Wave legal tech I define as technology that performs legal work, i.e. it actually conducts legal tasks, such using AI for reviewing a document, or guiding a client via an expert system through a complex legal issue.

We have never before been in this situation. We have had some automation, e.g. with document creation, but this was very much a manual, *PLUS lawyer, endeavour*. Once trained, or developed, the new range of AI legal tools are able to conduct elements of work by themselves. That is a new level of automation we have not seen before.

In turn this opens up questions about staffing of law firms, the role of paralegals, the pricing of legal services and how clients will respond.

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

They should be informed of what narrow AI is and what it can do. The Government has a responsibility to prepare the UK for the coming changes, with schools likely to need to introduce special modules on AI to prepare the future workforce.

The UK, and Western nations as a whole, may be about to face the greatest skills re-training challenge since the tail end of the Industrial Revolution. The reality is that a significant number of adults cannot see a great future ahead if low skilled roles are taken from them.

But, the idea of 'universal income' seems to be a very risky economic plan that would 'lock in' millions of people into poverty and hopelessness for generations. Far better to make a major investment in training, both for young people and retraining adults.

In sum: **the AI and automation revolution that is coming in turn will create the greatest ever demand for increased investment in education this country has ever seen.** (At least if we want to sustain a high quality of living and not become a split society like those such as in some developing nations, i.e. wealthy 'gated communities' of professionals vs under-employed communities with little hope of economic improvement.)

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

We are gaining huge increases in productivity, efficiency and the ability to do new things we could not have done before as a society because of AI.

We will all gain if those whose jobs are under 'skill erosion' can be retrained. The alternative is that those who own AI tech will accumulate great wealth and many others will see a dramatic decline in income. It is that stark.

The future does not have to be negative, but if left unplanned, then we do indeed face difficult times if for example, 30%-plus of the working population in are super-low skilled jobs, filling in the gaps that machines have left, and with the

percentage set to rise in the years ahead, unless training and new jobs are focused upon.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

(see above)

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

All sectors will benefit, see QU 4.

7. How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

(As above).

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

AI Ethical issues are surprisingly few, in my view. The main issue is algorithmic transparency so that we know what we have programmed software to do.

But, our choices as humans are already governed by law and ethics, therefore, the real question is ensuring that any law that covers human action is also 'followed across' into any AI systems we make.

Ironically, I don't see a special AI ethics problem in the law at all. This is and always will be a human problem first. I.e. a company that is biased in relation to gender, or social group, or ethnicity in terms of employment, is already breaking the law today in terms of its behaviour. Such a company may then make biased AI systems in the future, e.g. to help with employment decisions, but that company's culture is really the problem, not the AI system. We need to address any bias at the root of corporate culture, that is where it will 'leak' into any algorithms that companies design in the future.

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

We already have terrible levels of 'box ticking' and bias in human society. E.g. the approval of bank loans, for example, or credit ratings.

What happens with AI is no different and we need to fix the issues we have now, which are many.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Government can do very little to impact the pathways of technology. It is largely out of their hands. Banning certain types of algorithm seems legally impossible and unsustainable, especially in a world where technology is globally accessible.

What Government can do is to use taxpayers' money to fund new educational initiatives to ensure we don't create a two-tier society, of those who work *with* AI systems, and those who fill in the 'left over' super-low skill gaps.

One might say the mission is to avoid returning to Victorian times, where many millions of people lived in poverty, doing low skilled jobs, while an educated and high skilled professional and managerial class prospered.

There is a more socially fluid and equitable path forward that allows people to develop careers and generate wealth for themselves and the nation, across all regions and social groups, **but we all need the educational tools to achieve this.**

Government is also going to have to be both realistic and imaginative about how to help those over the age of 40, who will need to work for another 30 years in some cases, but may in the near future have little to no useful skills, other than for the above mentioned super-low skill jobs that generate very little productive benefit.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

Richard Tromans – Written evidence (AIC0227)

I see no particular initiatives there, other than ones focused on regulation, which seems unhelpful. The focus of global bodies needs to be on education and training.

Richard Tromans

Founder, ArtificialLawyer.com
Awarded 'Top 20 Best AI Sites in the World', 2017

Founder and Consultant
TromansConsulting

30 October 2017

UCL Knowledge Lab – Written evidence (AIC0105)

Evidence submitted by the Artificial Intelligence research group at UCL Knowledge Lab (KL): an interdisciplinary research centre at the IoE. UCL, IoE and the KL have an internationally leading reputation for excellence in education and research. We produce impact and innovations to enhance people's lives globally.

Executive Summary

- *AI is concerned with increasing our understanding about how the intelligent human mind works* by investigating the problem of designing machines that have intelligent abilities, as well as being about building these intelligent machines to address problems.
- Currently AI is somewhat technically biased towards Machine Learning (ML) with less attention being paid to the interdisciplinary nature of AI and the wider scope of what AI is about beyond ML.
- The advances in the *technology* of AI, in particular AI is worthy of excitement. However, there is a serious risk that public ignorance and fear will hamper progress. *Few if any of the potential benefits of AI, from personalized medicine to increased productivity through automation, will be achieved at scale unless we address the educational and training implications of AI now.*
- Education and training must equip people through investment in *carefully designed use of AI to improve education and help us address some of the big challenges we face. AND through educating people about AI, so that they can use and benefit from AI and so that they can actively contribute to the debates and developments within AI in an informed way. Implicit and primary in this question is the urgent requirement to educate the educators and trainers who will be expected to educate everyone else. Failure to address the educational and training implications of AI is likely to result in a failure to galvanize the prosperity that should accompany the AI revolution.*
- Companies who trade in data, such as Facebook, Amazon and Alphabet are generating enormous amounts of power because of their control of huge amounts of data. This data is not merely valuable because it comes in huge quantities, but also because of the way that these companies' process and refine it. The power manifested in these companies must surely be constrained in order to avoid the inevitable monopolization of the personal data market and its refinement.
- And yet in parallel, *collecting, collating and refining data, and extracting meaning from it is also the mainstay of much work within the communities of researchers who process educational data with the help of AI.* The latter community is *hampered by lack of investment* and by their adherence to ethical standards and protocols. It is right that any educational application of AI must be ethically designed and approved, it is also right that much

more investment should be made in the development of educational applications of AI. It is also right that *Large technology companies should be required to adhere to the same standards of ethical regulation as researchers when dealing with personal data.*

- There are very few genuine situations when the AI systems (as opposed to the data being processed and the results of that processing) should not be transparent about what the system is doing.
- Lack of transparency about the personal data that is collected from people *without their explicit informed* permission will undermine confidence in AI as inevitable misuses come to light. Clearly these *informed* permissions can only be achieved if people are educated about AI and are given the skills to influence its development. *As a society we need to empower individual members of the public to take charge of their personal data, we need to show them how to harness this data for their own benefit and give them the tools to scrutinize the algorithms or at least the decisions that these algorithms take.*
- The Government should take a pro-active role in the development and use of AI in the United Kingdom through the formation of a cross departmental interdisciplinary UK commission for AI to ensure that the growth of the AI sector is coordinated across stakeholders within and beyond the UK.
- *Human choices about how AI will be used in different settings will be the greatest decider of who benefits from AI and how. We must apply human intelligence judiciously to reap the fairest benefits.*

Definition of Artificial Intelligence (AI)

1. For the purposes of this submission, we use the Oxford dictionary definition, which defines AI as: *computer systems that have been designed to interact with the world through capabilities (for example, visual perception and speech recognition) and intelligent behaviours (for example, assessing the available information and then taking the most sensible action to achieve a stated goal) that we would think of as essentially human.* We also add that AI is an interdisciplinary area of study that includes psychology, philosophy, linguistics, computer science and neuroscience. The study of AI is complex and the disciplines are interlinked as we strive for a greater understanding of human intelligence as well as attempting to build smart computer technology that behaves intelligently. This definition is where the complexity of AI starts to unfold, because the definition itself relies upon an understanding of the term: intelligent behaviours, or more specifically of the word *intelligence*, an expression that is also the subject of multiple definitions. For example, is intelligence the ability to acquire and apply knowledge, is it wisdom, or is it the ability to handle criticism without blame or anxiety?

2. AI is concerned with increasing our understanding about how the intelligent human mind works by investigating the problem of designing machines that have intelligent abilities, as well as building these intelligent machines to address problems. The work of AI involves the combination of multiple disciplines, including: cognitive psychology to help us understand human

abilities such as problem solving, memory, vision and learning; philosophy of mind to shed light on what it is to be human, what consciousness is and why it is important for of our own intentional; computer science, mathematics, and logic to enable us to build complex technologies that can process information at speed; and linguistics to explain the structure and functions languages for communicating and thinking. The integration of so many different subject areas into the 'discipline' of AI, plus the breadth of activity that human intelligence affords are part of the reason that AI is so hard to define. We also need to contend with the difference between general intelligence and domain specific intelligence. This is translated in AI terms into the singularity: General AI - the singularity: the point at which an AI-powered computer or robot becomes capable of redesigning and improving itself or of designing AI more advanced than itself. This is general AI and it would have to successfully perform any intellectual task that a human being could perform. Domain specific Intelligence is much more limited and focuses on a single sort of intelligent activity. Domain specific AI is what all current AI does, such as playing games like chess and Go, recognizing people's faces and matching them to passport information, or driving a car.

Question 1. The pace of technological change: Current state of AI and contributory factors

3. The current state of Artificial Intelligence (AI) is somewhat technically biased towards Machine Learning (ML) with less attention being paid to the interdisciplinary nature of AI and the wider scope of what AI is about beyond ML. The factors that have contributed to this situation are most likely the demonstrable commercial and academic progress of ML from autonomous vehicles, through medical diagnosis and world champion standard 'Go' playing. This situation is likely to continue. However, the public ignorance about what AI is and what AI can and cannot do is likely to hinder progress unless well-designed interventions are made to ensure a much broader public understanding of AI. AI technology will likely continue to develop at a fast pace as large amounts of money are invested in commercially driven projects. For example, the Obama administration announced that it planned to invest US\$4 billion over a decade to make autonomous vehicles viable¹³²⁸, and recruitment of AI staff as grown rapidly: Amazon's average annual investment is \$227.8 million with 1178 AI jobs posted and Google \$130.1 million for average annual investment in its AI recruiting efforts - the company listed 563 AI jobs in the past year¹³²⁹. However, the societal change and public awareness will be much slower to develop and may well hinder progress in expected directions. We explain this further in Q3.

¹³²⁸ Spector, M. & Ramsey, M. U.S. proposes spending \$4 billion to encourage driverless cars. *The Wall Street Journal* (14 January 2016); <http://go.nature.com/2jZePEM>

¹³²⁹ <https://www.forbes.com/sites/aarontilley/2017/04/18/the-great-ai-recruitment-war-amazon-is-on-top-and-apple-is-almost-nowhere-to-be-seen/#60fa2bb361e5>

Question 2. The pace of technological change: Justification of current excitement

4. The developments in AI technology warrant excitement. There have been major developments in multidisciplinary implementations of AI and impressive technical progress in machine learning applications from the likes of Tesla, Google DeepMind, Amazon and IBM Watson. Excitement about the potential benefits and risks of AI are far less warranted with exaggerations frequently occurring in the media with descriptions of how AI is “mirroring how the human brain works.”¹³³⁰ Plus over optimistic AI rollouts falling over within hours.¹³³¹ There are also stark warnings from high profile experts.¹³³² There is a serious risk that without more attention being paid to ensuring that the general public understand enough about what AI is and is not, and about what AI can and cannot do, then ignorance and fear will hamper progress. We need more accessible and clearly written resources for the general reader that are based on evidence to support the claims that are made.

Question 3. Impact on society: Preparing the general public

5. Few if any of the potential AI benefits, from personalized medicine to increased productivity through automation, will be achieved at scale unless we address the educational and training implications of AI now.

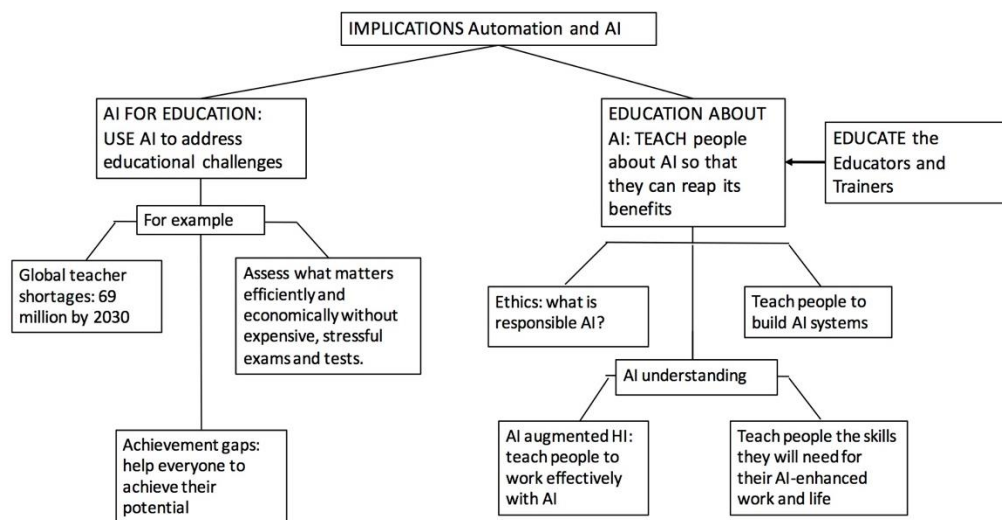


Figure 1: The AI and Education Knowledge Tree (Adapted from [1])

6. The nature of what needs to be done is illustrated in Figure 1. There are two key questions to be addressed:

A. How can we use AI to improve education and help us address some of the big challenges we face?

¹³³⁰ see for example, <https://www.theguardian.com/commentisfree/2017/mar/15/artificial-intelligence-deepmind-singularity-computers-match-humans>

¹³³¹ see for example, <http://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/>

¹³³² see for example, Stephen Hawking <http://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-could-wipe-out-humanity-when-it-gets-too-clever-as-humans-a6686496.html>

B. How can we educate people about AI, so that they can use and benefit from AI and so that they can actively contribute to the debates and developments within AI in an informed way? Implicit in this question is the urgent requirement to educate the educators and trainers who will be expected to educate others.

Question A. Addressing educational implications of AI with AI

7. The thoughtful design of AI approaches to educational challenges has the potential to provide significant benefits to educators, learners, parents and managers. But it must not start with the technology, it must start with a thorough exploration of the educational problem to be tackled. The development and adoption of AI teaching assistants¹³³³ will provide an opportunity for developing deeper teaching skills and enriching the teaching profession. This deepening of teacher expertise might be at the subject knowledge level, or it could be concerned with developing the requisite skills to support and nurture collaborative problem-solving in our students. It could also result in teachers developing the data science and learning science skills that enable them to gain greater insights from the increasingly available array of data about students' learning. However, whilst general funding for AI in the UK has multiplied, there is very little investment in AI for Education. This is immensely short-sighted when evidence shows that educational applications of AI can be extremely effective¹³³⁴.

Question B. Addressing educational implications of AI by educating about AI

8. There are three key elements that need to be introduced into *what we teach* as part of the curriculum across the sectors and in the workplace.
- The first is that *everyone needs to understand* enough about AI to be able to work with AI systems effectively so that AI and human intelligence (HI) augment each other effectively and we benefit from a symbiotic relationship between the two;
 - The second is that *everyone needs to be involved* in a discussion about what AI should and should not be designed to do. Some people need to be trained to tackle the legal and ethical aspects of AI in depth and help decision makers to make appropriate decisions about how AI impacts on the world;
 - Thirdly, *some people also need to know* enough about AI to build the next generation of AI systems.

There are also changes that need to be made to *how we teach* and train across the sectors and the workplace.

¹³³³ For example, as described here: <https://howwegettonext.com/a-i-is-the-new-t-a-in-the-classroom-dedbe5b99e9e> AND here: <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/Intelligence-Unleashed-Publication.pdf>

¹³³⁴ We outlined this in our submission to the House of Commons inquiry into Robotics and AI <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and-technology-committee/robotics-and-artificial-intelligence/written/32656.html>

9. *It is immensely short-sighted to continue to focus our education system on developing the routine cognitive skills that can most easily be automated.* A much more sensible and strategically valuable approach would be to focus on using teaching and training approaches that develop metacognition and self-efficacy: two concepts that are inter-linked and essential for lifelong learning. We also need to increase people's understanding across and between disciplines and their ability to work with others, both Human and Automated to solve real world problems. Collaborative problem-solving is a key skill for the workplace, and its importance is only likely to grow as further automation takes effect. There is currently a mismatch between the substantial evidence in favour of collaborative problem solving and learning reported in the literature and the approaches widely used within schools. This is neither preparing students for university nor the future workplace. For example, in an [interview](#) for a Davos 2016 debate on the Future of Education, a student from Hong Kong stated that the current school system produced: "industrialised mass-produced exam geniuses who excel in examinations" but who are "easily shattered when they face challenges". We need employees to be able to tackle challenges and this often involves working effectively with others to solve the problem at the heart of any challenge; we don't need exam geniuses who crumble under the pressure of the real world.

10. Collaborative problem-solving, just like other key skill developments, does not happen spontaneously. Both teachers and students require *a high level of training* to employ collaborative problem-solving effectively, and yet there is little evidence of concerted training effort. This might be due to lack of resources to provide such training and support. However, AI for education implementations have good potential to contribute to training students and teachers on their key skill developments, in addition to their routine cognitive abilities.

Implications for teacher training and professional development

11. The significant educational implications that AI brings to society, both when AI is viewed as a tool to enhance teaching and learning and when AI is viewed as a subject that must be addressed in the curriculum, make it clear that teacher training and teacher professional development must be reviewed and updated.

12. If teachers are to prepare young people for the new world of work, and if teachers are to prime and excite young people to engage with careers designing and building our future AI ecosystems, then we must train the teachers and teacher trainers and prepare them for their future workplace and its students' needs. This is a role for policy makers, in collaboration with the organisations who govern and manage the different teacher development systems and training protocols across countries. The need for young people to be equipped with a knowledge about AI is urgent, but the need for educators to be similarly equipped is critical.

Question 4. Impact on society: Who in society is gaining the most and the least

13. At the moment the big technology companies are gaining a great deal of commercial advantage from the combination of huge amounts of data and AI¹³³⁵. There are also significant developments in medicine and transport that ought to bring positive benefits to society. There is much debate about job losses and job gains due to AI in the workplace, but with little consensus and all we can be confident about is that employees will need to be flexible about their role in any organization, always willing and able to learn effectively. Therefore, any failure to recognize and address the urgent and critical teaching and training requirements precipitated by the advancement and growth of AI is likely to result in a failure to galvanize the prosperity that should accompany the AI revolution. In particular, we need to ensure that *everyone* is able to access the sort of education that will prepare them for a turbulent future workplace, not just those with the financial resources to access the limited amount of AI expertise in order to ensure that they are well prepared. As ever in times of rapid change, the poor and disadvantaged are likely to gain the least from AI. The major way to mitigate the potential disparities in the beneficial impact of AI on society is to address education and training (see Q3).

Question 5. Public perception

14. With the increase of public awareness of AI there are also growing concerns and evidence that despite a general optimism and hopes for positive uses, worries of loss of control over AI and potential negative effects continue to be discussed.¹³³⁶ Without a doubt great effort must be made to improve the public's understanding of, and engagement with AI by addressing the educational and training requirements as outlined in Question 3.

Question 6. What are the key sectors that do and do not stand to benefit from AI development?

15. At the moment the sectors most likely to be impacted on from AI are the sectors where large amounts of money have been invested in AI technologies. These are sectors where the ability to process large amounts of data to identify patterns and relationships are the core of what is required. Processing patient and treatment data for medical diagnosis, finding specific information from millions of documents in the legal profession or recognizing the identity of a person and their right to enter a country. It is hard to know who will benefit from these changes however. For example, more accurate diagnosis and treatment planning should benefit patients, but if their access to humans who can help them understand their disease and its treatment, empathize with their reactions and concerns, not least their emotional reactions, then, the patients' benefits from the automated diagnosis and treatment planning may be reduced. *Human choices about how AI will be used in different settings will be the greatest*

¹³³⁵ <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>

¹³³⁶ See for example, FAST, E.; HORVITZ, E. Long-Term Trends in the Public Perception of Artificial Intelligence. AAAI Conference on Artificial Intelligence, North America, Feb. 2017. Available at: <<https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14581>>

decider of who benefits from AI and how. Those with the least voice in society are least likely to be able to access the benefits, but human decisions could alter this situation.

Question 7. How can the data-based monopolies and the ‘winner takes- all’ economies be addressed?

16. In May 2017, the Economist¹³³⁷, suggested that “data is the oil of the digital era”. The article drew attention to the fact that the large technology companies who trade in data, such as Facebook, Amazon and Alphabet are generating enormous amounts of power because of their control of huge amounts of data. Data about almost everyone in the digitally enabled world. This data is not merely valuable because it comes in huge quantities, but because of the way these companies’ process and refine it: to tell them what people are buying, what they are searching for and who they are connecting with.

17. The power manifested in these companies must surely be addressed and reduced in order to avoid the inevitable monopolization of the personal data market and its refinement. And yet, collecting, collating and refining data, and extracting meaning from it is also the mainstay of much work within the communities of researchers who process educational data with the help of AI. The latter community is hampered by lack of investment and by their adherence to ethical standards and protocols. It is right that any educational application of AI must be ethically designed and approved, it is also right that much more investment should be made in the development of educational applications of AI. And it is right that the large technology companies should be required to adhere to the same standards of ethical regulation as research scientists.

Question 8. What are the ethical implications of the development and use of artificial intelligence?

18. The potential for misuse of both personal data and the algorithms that process this data is huge. The lack of transparency about the personal data that is collected from people without their explicit *informed* permission will undermine confidence in AI as inevitable misuses come to light. Clearly these *informed* permissions can only be achieved if the people are educated about AI and are given skills of how to influence its development as explained in question B of the question 3 above. Another concern is the potential for bias (conscious or unconscious) to be incorporated into AI. As a society we need to empower individual members of the public to take charge of their personal data, we need to show them how to harness this data for their own benefit and give them the tools to scrutinize the algorithms or at least the decisions that these algorithms take. The way in which AI is used to process people’s personal data must also be subjected to regulation to ensure that it is fair and transparent about what the processing is designed to achieve, even if the detail of how the processing is completed remain private for commercial reasons.

¹³³⁷ <https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>

Question 9. In what situations is a relative lack of transparency in AI systems acceptable?

19. There are very few genuine situations when the AI systems (as opposed to the data being processed and the results of the processing) should not be transparent about what the system is doing. There is a much stronger argument for 'black boxing' about how an AI system is achieving the processing and results, which may be commercially sensitive. The main cases when 'black boxing' is justified are concerned with security, but even in this sector there is a risk of overstating the need for a lack of transparency. The vast majority of data about people and the results that AI produces from processing this personal data should be under the control of the person whose data is being used (or their parent/guardian in the case of a minor). Each individual can then make a decision about what they wish to share and with whom. The implications for a system based on personal data ownerships such as this include the need for individuals to understand enough about their data to take responsibility for it. Such transparency is also key to improving people's understanding of AI systems.

Question 10. The role of the Government:

20. The Government should take a pro-active role in the development and use of AI in the United Kingdom through the formation of a cross departmental interdisciplinary UK council for AI to ensure that the growth of the AI sector is coordinated and promoted within and beyond the UK across stakeholders.

21. The Government must pay specific attention to the needs of those with the least voice in society in order to ensure that they too can access the benefits of AI. As stated in answer to Q6, human decisions about how AI is to be developed will be the deciding factor in who benefits from AI: Government must lead the way.

Question 11. Learning from Others: What lessons can be learnt from other countries

22. There are many other countries who are setting examples from which we can learn. For example:

In China, Beijing Municipal Commission of Education launched the "Advanced Innovation Center Construction Plan of Higher Education in Beijing" to "integrate national, domestic and international resources, to promote both research and application, to combine technological creation and talent development, to develop both national and local colleges and universities". The new centre at Beijing Normal University has a remit to conduct research in AI through their AITutor project drive an AI transformation of Beijing public education¹³³⁸.

23. Finland, one of the leading countries for a successful education system, as evidenced in their OECD PISA rankings, is significantly reforming its education

¹³³⁸ <http://aic-fe.bnu.edu.cn/en/about/index.html>.

system¹³³⁹. They are revising both what they teach and how they teach it. The country's education committee plans to change the curriculum of what is taught in order to prepare students better for their work life in an automated AI enhanced workplace. In addition, students will learn better communication skills when working in collaboration with their classmates (note answers to Q3).

6 September 2017

¹³³⁹ https://ec.europa.eu/education/sites/education/files/monitor2016-fi_en.pdf

UK Computing Research Committee – Written evidence (AIC0030)

Response to the Call for Evidence by the House of Lords Select Committee on Artificial Intelligence

Compiled on behalf of the UK Computing Research Committee, UKCRC.

Coordinated by:

Chris Johnson
Professor and Head of Computing Science,
School of Computing Science, University of Glasgow, Glasgow, G12 8RZ.

UKCRC is an Expert Panel of the British Computer Society (BCS), the Institution of Engineering and Technology (IET), and the Council of Professors and Heads of Computing (CPHC). It was formed in November 2000 as a policy committee for computing research in the UK. Members of UKCRC are leading computing researchers who each have an established international reputation in computing. Our response thus covers UK research in computing, which is internationally strong and vigorous, and a major national asset. This response has been prepared after a widespread consultation amongst the membership of UKCRC and, as such, is an independent response on behalf of UKCRC and does not necessarily reflect the official opinion or position of the BCS or the IET.

Questions

The pace of technological change

1. What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

It is likely that techniques, which are today collectively labelled as 'artificial intelligence' (AI) or machine learning (ML), will become more commonplace within a wide range of computational and embedded systems. It seems likely that this may intersect with other developments in computing, including the Internet of Things and Smart Cities.

These developments raise considerable challenges – especially in terms of the interactions that arise when AI applications make inferences about human behaviour and vice versa. The practical impact of this is being seen in US states where human drivers can now take additional driving lessons on how to avoid accidents with autonomous vehicles.

Another key area is the regulation of AI related systems, for instance in safety critical systems. It is hard to demonstrate the safety of algorithms that may evolve or learn over time or when training sets cannot match all of the possible environmental situations that an application might meet. These issues are visible now in the evolving regulations applied to autonomous vehicles but this is a more general concern.

2. Is the current level of excitement, which surrounds artificial intelligence warranted?

Yes – although there is some hype that exaggerates what is possible in the immediate future. There is a need to distinguish between areas where there is realistic prospect of revolutionary changes in the next 10-20 years and areas where changes will be much slower (e.g., because of poor quality data or the lack of tractable algorithms for addressing recognised problems).

Impact on society

3. How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

This is part of a far wider question about the need to prepare society for future developments within information technology and networked systems. The UK lags behind many other states in terms of the attention paid to the teaching of Computing Science (as opposed to IT-training which focuses on the ability to use particular applications). Specific areas of government are doing their best to address this concern – for example the NCSC initiatives in cyber education for schools. Initiatives to improve computing science education in the UK are poorly coordinated. They are isolated in silos that result from the particular focus of individual government departments.

The biggest impact of AI will be on the future of work. It will affect when, where and how people engage with computing technologies. We will see a declining importance of some skills sets and a rise in others. It is likely that the skills required for routine knowledge-based work will decline in value, while those

dealing with exceptional cases will rise in value. There will be a particular need for strong social skills and human negotiation to resolve these exceptional cases. We should engage the population in more informed discourse on that nature and value of data privacy, balanced against the value of data sharing (particularly in domains such as healthcare).

As well as preparing the general public, Government must itself be prepared for what looks like the biggest disruption since the Industrial Revolution. Automation, fuelled by new technologies including AI, looks set to undermine many assumptions in society concerning people's everyday lives: jobs, education and training, but also remuneration, and leisure.

4. Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

Many UK companies now use large-scale data analysis techniques, which would previously have been termed 'artificial intelligence'. This trend is likely to continue – for instance, the use of fuzzy reasoning within embedded devices such as the variable speed controllers of washing machines. In most cases, users are unaware that these embedded systems use AI algorithms.

In terms of UK research, it is possible to identify a cluster of companies that fund and then exploit University projects. Many are US based – in particular, Google, Amazon, Microsoft. This reflects market dominance within the software industry and may also illustrate a need to focus support for UK industry in this area.

There is a risk that developments favour the privileged and further disadvantage those with lower digital literacy; they may also favour larger organisations, at the cost of smaller organisations (e.g. those in the voluntary / charity sector) that do not have the capacity to exploit the new capabilities.

A first step to mitigating the risks of greater disparities is an increasing focus on technology education – not just through formal education, but life-long learning, so that people of various ages and backgrounds are empowered to engage with developments.

Public perception

5. Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Yes – as part of a wider and coordinated programme to improve the teaching of Computing Science in UK schools. There is a lack of scientific research into the

pedagogy of computing – we should identify effective ways of teaching the topic and engaging especially with under-represented groups as a means of addressing the gender and racial biases that propagate into University. This should also extend beyond formal education into life-long learning so as to be inclusive of older people.

Industry

6. What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

This is a very broad question – all sectors have potential to gain through the application of AI and ML to data analysis. The public sector could do more to benefit from these techniques to support the provision and optimisation of services across a host of areas related to urban planning, healthcare etc. Transport is already making big steps towards the application of control-based algorithms for autonomous vehicles but the regulatory issues mentioned earlier are a significant concern.

More broadly, sectors where quantification is valuable, and where there are existing or potential large bodies of data, stand to benefit. Those that depend more on “soft skills” that are not computationally tractable are less likely to benefit significantly. It is important that, with the growing focus on artificial intelligence, society forgets to value natural intelligence too.

7. How can the data-based monopolies of some large corporations, and the ‘winner- takes-all’ economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

Data protection laws place a limit on the disclosure of information but there is a lot to be gained through the provision of APIs or interfaces to aggregate data held by the large corporations so that we can develop an ecosystem of SMEs – archetypal app developers, to generate a more vibrant UK ecosystem in this area.

Ethics

8. What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

Studies of the combination of ethics and law should be funded; especially where AI will be used in critical systems. Particular concerns focus on the application of

AI in health, transport (see also in 10 below), and also in security and resilience mechanisms designed for use in Critical Infrastructures Protection such as Smart Grids.

As a consequence, the law needs to be updated. Legal and ethical experts need to be educated, preferably in studies combined with technology (see also in 3 above). For example, questions of liability arise when human road-users are in collision with autonomous vehicles. Would there be a degree of culpability associated with the operators of the autonomous vehicle and with the engineers who coded or tested the AI application?

9. In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

The laws protecting intellectual property provide well-framed principles for transparency in software engineering; as do the existing regulatory provisions in safety and security critical systems. Problems arise when it is hard to apply existing techniques to determine the reliability of these systems because of the characteristics of AI and machine learning algorithms. These systems, typically, generalise from learning sets to influence behaviour when faced with previously unseen environments. Such approaches undermine existing regulatory provision unless we can require exhaustive testing to help ensure appropriate responses across potential operating environments (this approach is being developed by US National Highways and Transport Safety Agency for approval of autonomous vehicles, UK CRC also supports the team in the Dept for Transport working on connected and autonomous vehicles). Exhaustive testing had previously been widely rejected as an acceptable basis for the engineering of safety and security related systems – how can we be sure that all future behaviours have been considered across millions of lines of code. The resolution of these tensions remains a topic of active research; even having such transparency can provide few guarantees for regulators or the UK public. Related issues include the use of learning – where the behaviour of AI/ML can change over time as new training sets are used – creating non-determinism; hence the behaviour seen in previous environments may not be a reliable guide to future performance.

The role of the Government

10. What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

AI and ML are algorithmic technologies. Regulation must focus on the application of these approaches. These applications extend across many different branches of government – with autonomous technology being applied in power network

management, healthcare, transport etc. There is a strong need to commission studies that identify appropriate regulatory mechanisms that are consistent between these areas. For example, transport will most certainly need new bodies to set up and develop regulation of driverless vehicles including cars, trucks, buses, trams and trains as well as in the aviation industry.

Learning from others

11. What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

As mentioned above, the US National Highways and Transport Safety Agency has been innovative in both promoting the use of AI in autonomous vehicles but also in ensuring safeguards. Their use of waivers to permit testing that might otherwise violate federal law is quite different and arguably not so useful as the UK guidelines denoting 'best practice'. However, the NHTSA move to enable the reclassification of certain AI algorithms as the driver of the car is innovative as is their approach to testing. One major caveat here is the lack of access to the data being generated by companies through the testing – to improve public confidence that they are not being placed at risk by these tests.

28 August 2017

The Association for UK Interactive Entertainment (Ukie) – Written evidence (AIC0116)

Introduction

UK Interactive Entertainment (Ukie) welcomes the opportunity to respond to this inquiry. Our response is intended to provide the Committee with a robust understanding of the role the games industry is playing in the development of artificial intelligence (“AI”), as well as the innovative and creative uses of AI pioneered by the games industry, and why our sector stands to benefit from its continuing development. We also explore how the games industry can play a role in helping improve the public’s engagement with AI, and understanding of the innovative new experiences and tools for creativity it presents. Finally, we outline the creative and computational skills we believe will be fundamental to the growth of the UK’s economy, particularly in light of more widespread use of AI, and what role the government can play in supporting the development of these skills.

About Ukie

Ukie (UK Interactive Entertainment) is the trade body for the UK's games and interactive entertainment industry. A not-for-profit, it represents games businesses of all sizes from small start-ups to large multinational developers, publishers and service companies, working across online, mobile apps, consoles, PC, eSports, Virtual Reality and Augmented Reality.

Ukie aims to support, grow and promote member businesses and the wider UK games and interactive entertainment industry by optimising the economic, cultural, political and social environment needed for businesses to thrive.

About the UK games industry

The UK games and interactive entertainment industry is an international success story, with the potential to take an ever-larger export share of a global market that will soon be worth more than \$99 billion.¹³⁴⁰ The UK is already well positioned as a significant player in this field and is currently estimated to be the sixth largest video games market in terms of consumer revenues, with an estimated worth of £4.33bn.¹³⁴¹

The UK games industry blends the best of British technological innovation and creativity, resulting in successful games and technology which are exported around the world and which cross over into other creative sectors. By way of illustration, *Grand Theft Auto V*, the biggest-selling entertainment product of all time (generating \$1 billion in global revenues in just three days following its

¹³⁴⁰ <https://newzoo.com/insights/articles/global-games-market-reaches-99-6-billion-2016-mobile-generating-37/>

¹³⁴¹ <http://ukie.org.uk/research#Market>

release), and ground-breaking video games, such as the *Batman Arkham* trilogy, *No Man's Sky* and *Elite* are the brainchildren of UK developers.

The games industry is also playing a leading role in the development of emerging technologies such as AI as well as big data analysis, virtual reality and augmented reality which are each expected to drive high value growth markets in the games industry as well as other sectors like health and education.

The pace of technological change

Artificial Intelligence & the games industry

This section addresses elements of questions 1, 4 and 6 of the Committee's questions. It will firstly set out the important role that the games industry has played in the research and development of AI, and secondly highlight the innovative and creative ways that AI has been used by the games and interactive entertainment industry to create compelling new experiences for players. It will conclude by explaining how the industry stands to benefit from its ongoing development.

The games industry's role in the development of artificial intelligence

In March 2016, a historic milestone for AI was reached when the Google DeepMind's program, AlphaGo, defeated the world-class Go champion Lee Sedol in the ancient board game with more possible moves than atoms in the universe.¹³⁴² This advancement rightly garnered significant global media attention,¹³⁴³ and highlighted the important role that games play in the development of AI.

Since as early as 1949, when Claude Shannon published his thoughts on how a computer might be made to play Chess¹³⁴⁴ and 1951 when Alan Turing published his famous algorithm TurboChamp¹³⁴⁵, computer scientists have been using games as an effective tool to measure how good a computer can become at performing specific tasks that challenge the human intellect.

The AI community has made it very clear that they view videogames as the best platform to use to advance AI. In the last twelve months, arguably the two biggest AI research companies in the world - Google's UK-based DeepMind, and Elon Musk's OpenAI - have both made important commitments to using videogames as the main platform for their research. DeepMind is using Atari

¹³⁴² <https://research.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html>

¹³⁴³ <http://www.theguardian.com/technology/2016/mar/09/google-deepmind-alphago-ai-defeats-human-lee-sedol-first-game-go-contest>

¹³⁴⁴ <http://www.andreykurenkov.com/writing/a-brief-history-of-game-ai/>

¹³⁴⁵ <http://www.andreykurenkov.com/writing/a-brief-history-of-game-ai/>

games as the primary test case for their deep learning research, and recently announced they are partnering with Activision Blizzard to build AI for *Starcraft 2*. In August 2017, OpenAI announced they are partnering with Valve Software to build an AI for *DOTA 2*. In both cases, videogames are seen as a rich and complex environment for AI to tackle, while still being a controllable environment and providing a huge amount of feedback.

AI researchers are continuing to find games to be an invaluable tool for a number of reasons. Firstly, games can provide AI a safe training ground to gather data which can then be used and adapted to the real world. By way of illustration, last year, AI researcher Artur Filipowicz of Princeton University discovered that the immensely popular – UK developed - game *Grand Theft Auto V* could be used to help develop an appropriate algorithm for autonomous vehicles to recognise stop signs. By making small modifications to the game, he was able to develop software that could navigate the traffic and read stop signs. *Grand*

Theft Auto V has winding city streets, mountains, and highways that can be explored in 257 different cars through 14 different weather simulations, making it an ideal simulated test-driving range for autonomous cars. This discovery subsequently led to OpenAI, in partnership with the DeepDive Project, releasing an open-source integration that enables *Grand Theft Auto V* to be used as a driving simulator for autonomous vehicles software, thereby notably accelerating the development of self-driving vehicle technology and making it cheaper, more accessible and safer than test driving autonomous vehicles on physical roads.¹³⁴⁶

Secondly, games offer researchers repeatable and safe learning environments which help machines improve their learning skills. For instance, DeepMind exposed its AI agent to Atari games without first teaching it how to play these games. The DeepMind program was eventually able to master all the Atari games it played, demonstrating how the repeatable and controlled environment of video games can enable AI agents to learn on their own.¹³⁴⁷

Thirdly, because different games require different cognitive skills, numerous AI researchers believe that games play a crucial role in helping them understand how the problem of intelligence can be broken down into smaller, more manageable chunks, and could potentially even help to develop a proper AI theory.¹³⁴⁸ By exposing its agent to Atari games and identifying which ones it found harder to master, DeepMind researchers were able to determine what

¹³⁴⁶ <https://www.inverse.com/article/26307-grand-theft-auto-open-ai>

¹³⁴⁷ <https://deepmind.com/research/publications/playing-atari-deep-reinforcement-learning/>

¹³⁴⁸ <https://www.economist.com/news/science-and-technology/21721890-games-help-them-understand-reality-why-ai-researchers-video-games>

tasks their agent struggled to achieve and improve their algorithms accordingly. They published research into how, by understanding why their agent failed at the Atari game *Montezuma's Revenge*, they could adapt its agent to be more curious, thereby enabling it to become more likely to develop good problem-solving strategies.¹³⁴⁹ The advancements made in this area by DeepMind were not only confined to mastering skills in a virtual world, but have been used to solve real world problems like reducing energy usage in Google's data centres by 40%.¹³⁵⁰

By providing an ideal training ground for the real world, games and game technology have been invaluable tools for AI researchers to test and improve their systems, and the games industry therefore stands to continue to play a significant role in the future development of AI.

The use and development of artificial intelligence within the games industry

The games and interactive entertainment industry has not only provided valuable tools for AI researchers from other fields, but is itself a sector which has significantly benefitted from developing and using AI as a creative tool to continuously create innovative, engaging and high-quality experiences for its consumers.

AI is fundamental to bringing virtual worlds to life and determining the way a player interacts with a game. It has been used as a tool in the games industry since not long after the origins of video games, where it was initially designed for creating non-player opponents in classic arcade games like *Pong* and *PacMan*.¹³⁵¹ Games companies continue to research and push the boundaries in creating more realistic, human-like opponents and companions for video games. For example, EA's SEED team recently developed a goal-based multi-action AI character that learns how to play a video game from using only visual and audio inputs that a human would have playing a game.¹³⁵²

As games have grown increasingly sophisticated, AI has been used to make games more entertaining and challenging to players, by allowing games developers to build engaging non-player characters (NPCs) and model the way NPCs interact, to simulate events taking place within games, as well as discerning the emotional state of a player and tailoring the game appropriately. A notable example of how AI enables games developers to create more compelling and rich experiences for players is the smash hit game franchise, *The Sims* which

¹³⁴⁹ <https://deepmind.com/research/publications/unifying-count-based-exploration-and-intrinsic-motivation/>

¹³⁵⁰ <https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/>

¹³⁵¹ https://sites.google.com/site/myangelcafe/articles/history_ai

¹³⁵² <https://twitter.com/seed/status/894708178289602561>

provides players with a household of intelligent characters who form relations and develop behaviours that emulate emotional depth and authenticity.¹³⁵³

The games industry provides a powerful example of a sector where advancements in technology have continuously enabled and fuelled the development of new forms of expressions and creativity. Similarly to how advances in motion sensing and capture technology spurred the development of a new generation of gaming systems, such as Microsoft's Kinect and Nintendo's Wii, operated through body movements rather than controllers, developments in AI are empowering games developers to create more compelling and realistic characters and richer worlds for their players.

Moreover, it is envisaged that developments in conversational interfaces, powered by AI, will change the way we interact with video games. For example, instead of a game using a dialogue menu system players will be able to use words to interact with non-player characters (NPCs). Thereby making NPCs feel more lifelike and helping to build more meaningful relationships between game characters and humans. Conversational interfaces could also help interact with games while offline (e.g. by telling a voice assistant to auction off certain in-game items without even entering the game).

Advances in AI are not only impacting characters inhabiting virtual worlds, but every aspect of game development. For example, procedural generation - the use of algorithms to create part of a game - has led to the development of huge games like *Minecraft* and *No Man's Sky*, which can create seemingly endless bespoke worlds while the player is playing. The UK developed game *No Man's Sky*, demonstrates the innovative and imaginative player experiences that can be created through using AI, by placing players in the role of an astronaut exploring a cosmos made of 18 quintillion procedurally generated life-size planets which each feature their own life, ecology, lakes, caves and canyons.¹³⁵⁴

Developments in AI have consistently provided games businesses with new tools to experiment and innovate with. As a sector we therefore stand to benefit from its continued use and advancement.

The future of artificial intelligence in the games industry over the next decade
Developments in AI are also impacting the way games are designed and produced. Massive open world games, like *No Man's Sky*, would have traditionally required large teams of developers (and associated development budgets) to design and draw every element of their game, but now, by using AI,

¹³⁵³ <https://www.theguardian.com/technology/2016/oct/12/video-game-characters-emotional-ai-developers>

¹³⁵⁴ <https://www.theguardian.com/technology/2015/jul/12/no-mans-sky-18-quintillion-planets-hello-games>

such games can be produced by much smaller teams of game developers. This provides new creative development opportunities for start-ups and independent games studios and helps further democratise games development.

Advances in AI are also beginning to change the way game developers are working. By way of illustration, at the 2017 London Games Festival's AI Summit, Imre Jele - co-founder of UK-based Bossa Studios, makers of the best-selling game *Surgeon Simulator* - gave a talk titled "Your Next Hire Should Be An AI", where he explained how Bossa Studios is using generative AI algorithms to contribute ideas and art assets to their art team. This highlighted how AI could bring significant changes to the way games companies are run, the scale of what they can achieve, and the way they contribute to the UK economy.

Moreover, as the preceding example demonstrates, these advances in AI could potentially transform the role computers play in the games development process: from simple tools employed by developers to becoming genuine collaborators in the creative process. Julian Togelius, a Professor at New York University's School of Engineering and AI consultant at British-based games startup Spirit AI, explains how game engines could use procedural generation, AI, and creative computing techniques to dynamically build environments and experiences to suit every individual player's unique desires.¹³⁵⁵ Similarly, AI developments could enable non-player characters to themselves generate new stories and dialogue based on player preferences entirely unique to that player's individual experience. In both examples, the traditional games design process is altered as the role of the games designer becomes to design a set of rules which vests creative power in AI to then invent and develop experiences for players itself.

These recent and forthcoming developments help to convey the significant creative potential of AI developed and applied in the games industry both in terms of empowering game developers to explore wholly new ways of creating games, and by offering players innovative interactive entertainment experiences that are uniquely relevant to them.

Public perception

The way new technologies and their risks and benefits are presented can markedly influence their development, regulation and place in public opinion. We believe that the games industry can play an important role in both helping to improve the public's engagement and technical understanding of AI, as well as foster a positive perception of AI and the innovative new experiences and tools for creativity it offers to players and creators alike.

¹³⁵⁵ <http://pcgbook.com/>

A major barrier to the widespread adoption of AI across society is overcoming some of the misconceptions and fears people might have, which can often come from a lack of everyday experience in interacting with intelligent software. Games provide a unique opportunity to prepare society for the future, offering them a safe space where people can take risks, make mistakes and be curious about AI.

Unlike other playful AI tools, like photo filters on smartphones, games are a two-way immersive interactive entertainment experience resulting in players interacting with an AI and seeing the results. They are, therefore, a great medium for people to learn how to interact and engage with AI systems, which is an important area of research for AI, especially in enabling greater safety of systems. There is already a wide range of games driven by their AI¹³⁵⁶, such as *Black And White* (in which players train a machine learning system, under the guise of training an animal) or *Alien: Isolation* (in which players must understand the strengths and weaknesses of an AI, represented as a dangerous alien). Both of these milestone AI-based games were developed by UK game developers.

Games also serve as an effective medium to capture the public's imagination and drive enthusiasm for the creative potential and uses of AI. *No Man's Sky*, for example, offers a vivid depiction of the artistic possibilities opened up by AI, by providing players the opportunity to marvel and explore beautiful and expansive planets, in which every rock, flower, tree, creature and scene is generated by an AI algorithm.

Moreover, by acting as an example for the ways AI can empower and fuel human creativity, the games industry can be used to help develop the public's understanding of the various opportunities that exist for society in the development and use of AI. AI presents huge potential to unlock individual creativity in areas that traditionally have a high barrier to entry. The creation of videogames is a good example of this, as it requires many artistic and technical skills to create even a simple game. Over the next decade, academics predict that we will see the emergence of 'computationally creative' AI systems that can tackle highly creative problems, which historically have been problematic for AI. Dr. Michael Cook at Falmouth University has done work in this area that vividly demonstrate this. - His ANGELINA¹³⁵⁷ system has created videogames on its own as well as in conjunction with humans, and is designed to be able to explain its actions, understand cultural references and common knowledge, and be inventive and novel.

Creative AI that can work with people and converse with them about creative tasks could change everyday creative expression, making it easier and increasing

¹³⁵⁶ <http://julian.togelius.com/Treanor2015AIBased.pdf>

¹³⁵⁷ <http://www.gamesbyangelina.org/>

everyone's potential for creating and sharing things like videogames. Increasing the public's understanding of the creative opportunities presented by the development of AI is important to foster an informed and balanced perspective on how AI will impact society.

Impact on society and the role of the Government

The development and widespread uses of AI across all sectors of the UK economy will continue to significantly increase demand for individuals with critical, creative, computational, and problem-solving skills. There is a significant role for the government to invest in the digital and creative skills needed to support a strong UK economy, especially in light of the well documented digital skills gap.¹³⁵⁸ This can best be achieved by focusing on the education curriculum, teacher training and digital inclusion. For example, whilst a renewed focus on coding in the curriculum is very welcome, it is important that teachers are fully trained in how to deliver it, and that government supports their training.

Students need to be prepared for a future where robotics and AI are commonplace, and our education system should be developing the cognitive skills that are not easy to automate. By way of illustration, if AI is widely used in games development as a tool to create personalised content and experiences for players, as described above, the skills required of certain games developers could evolve; they will need new skills to successfully design sets of rules and instructions from which an AI can help create a game.

This inevitably calls for a deeper understanding of computational and systems thinking, as games designers would in essence be designing sets of rules through which creativity can arise. Therefore, to prepare the future workforce for the widespread use of AI, they need to be equipped with the computational thinking skills necessary to conceive and design systems, as well as the creative skills to manipulate technology to deliver an innovative outcome. Supporting the development of creative skills alongside technical ones is crucial as innovation inherently relies on artistic and creative thinking.

In recent years Ukie has been at the forefront of advocating for changes to the UK's educational system to ensure that the creative, computational and critical thinking skills needed for the future growth of the UK's economy are properly embedded in schools and classrooms.¹³⁵⁹ The Ukie-led teach training program Digital Schoolhouse empowers, supports and trains teachers in their delivery of the computer science curriculum by providing creative workshops where both teachers and pupils learn about computing fundamentals through play-based learning techniques. Our national programme has established over 20 Digital

¹³⁵⁸ We're Just Not Doing Enough - Working Together to meet the Digital Skills Challenge, Tech UK 2015

¹³⁵⁹ <http://ukie.org.uk/content/next-gen-skills-campaign-launched>

Schoolhouses across the UK, which collectively supports almost 2000 teachers and over 10,000 pupils each year.

Conclusion

To conclude, we hope our response has highlighted to the Committee: (i) the important role the games industry is playing in the development of artificial intelligence, (ii) the creative ways AI is being harnessed by our sector to create innovative and engaging interactive entertainment, and (iii) how games can be used as a tool to increase the public's understanding and engagement with AI. Whilst, the games industry stands to benefit from the continued development and use of AI, we believe that there is a clear role for government to further support the development of the critical, creative and computation skills that will be vital to developing and using AI as well as equipping our workforce with the skills needed for the future growth of the UK economy.

6 September 2017

Dr Ozlem Ulgen – Written evidence (AIC0112)

The ethical implications of developing and using artificial intelligence and robotics in the civilian and military spheres

Summary

Machine-mediated human interaction challenges the philosophical basis of human existence and ethical conduct. Aside from technical challenges of ensuring ethical conduct in artificial intelligence and robotics, there are moral questions about the desirability of replacing human functions and the human mind with such technology. How will artificial intelligence and robotics engage in moral reasoning in order to act ethically? Is there a need for a new set of moral rules? What happens to human interaction when it is mediated by technology? Should such technology be used to end human life? Who bears responsibility for wrongdoing or harmful conduct by artificial intelligence and robotics? This paper seeks to address some ethical issues surrounding the development and use of artificial intelligence and robotics in the civilian and military spheres. It explores the implications of fully autonomous and human-machine rule-generating approaches, the difference between “human will” and “machine will, and between machine logic and human judgment.

About the author

Dr Ozlem Ulgen is a Visiting Fellow at the Lauterpacht Centre for International Law, and Wolfson College, University of Cambridge. She is Senior Lecturer in Law at the School of Law, Birmingham City University, and a barrister (non-practicing) called to the Bar in England and Wales. She specialises in moral and legal philosophy, public international law, and international humanitarian law. Her areas of expertise relate to cosmopolitan ethics in warfare, Kantian ethics and human dignity, and the law and ethics of autonomous weapons. She is currently writing a Routledge-commissioned monograph, *The Law and Ethics of Autonomous Weapons: A Cosmopolitan Perspective*. Forthcoming publications: ‘*World Community Interest’ approach to interim measures on ‘robot weapons’: revisiting the Nuclear Test Cases* (New Zealand Yearbook of International Law); *Pre-deployment common law duty of care and Article 36 obligations in relation to autonomous weapons: interface between domestic law and international humanitarian law?* (The Military Law and the Law of War Review); *Human dignity in an age of autonomous weapons: are we in danger of losing an ‘elementary consideration of humanity’?* (OUP edited collection).

Introduction

Artificial intelligence and robotics is pervasive in daily life and set to expand to new levels potentially replacing human decision-making and action.

Self-driving cars, home and healthcare robots, and autonomous weapons are some examples. A distinction appears to be emerging between potentially benevolent civilian uses of the technology (e.g. unmanned aerial vehicles delivering medicines), and potentially malevolent military uses (e.g. lethal autonomous weapons killing human combatants). Machine-mediated human interaction challenges the philosophical basis of human existence and ethical conduct. Aside from technical challenges of ensuring ethical conduct in artificial intelligence and robotics, there are moral questions about the desirability of replacing human functions and the human mind with such technology. How will artificial intelligence and robotics engage in moral reasoning in order to act ethically? Is there a need for a new set of moral rules? What happens to human interaction when it is mediated by technology? Should such technology be used to end human life? Who bears responsibility for wrongdoing or harmful conduct by artificial intelligence and robotics?

This paper seeks to address some ethical issues surrounding the development and use of artificial intelligence and robotics in the civilian and military spheres. It explores the implications of fully autonomous and human-machine rule-generating approaches, the difference between “human will” and “machine will, and between machine logic and human judgment.

Fully autonomous and human-machine rule-generating approaches

Artificial intelligence and robotics do not possess human rational thinking capacity or a free will to be able to understand what constitutes a rule that is inherently desirable, doable, and valuable for it to be capable of universalisation. But there is human agency in the design, development, testing, and deployment of such technology so that responsibility for implementing moral rules resides with humans. Humans determine which rules are programmed into the technology to ensure ethical use and moral conduct. For these rules to be capable of universalisation they must be “public and shareable”.

In the civilian sphere, for example, there is much debate about open access and use of artificial intelligence to gather personal data, potentially compromising privacy. In the military sphere, discussions on lethal autonomous weapons under the auspices of the UN Convention on Certain Conventional Weapons represent a process for universalisation of rules which may regulate or ban such weapons. Indeed, there is emerging *opinio juris* among some states for a preventative prohibition rule, and a majority of states recognise that any rules regulating lethal autonomous weapons must take account of ethical, legal, and humanitarian considerations.¹³⁶⁰

¹³⁶⁰ See, O Ulgen, “World Community Interest’ approach to interim measures on ‘robot weapons’: revisiting the *Nuclear Test Cases*” (2016) 14 *New Zealand Yearbook of International Law* (forthcoming) Section III.A.

The potentially broad purposes and uses of artificial intelligence and robotics technology may lead to competing rules emerging which may or may not be capable of universalisation. Some preliminary issues related to the nature and type of rules are considered here.

How will rules be generated to regulate ethical use and operation of the technology? This depends on whether the technology is intended to completely replace human functions and rational thinking or to complement and supplement such human characteristics. Fully autonomous technology refers to artificial intelligence and robotics replacing human rational thinking capacity and free will so that rules emerge from the technology itself rather than humans. Human-machine integrated technology, on the other hand, refers to technology that supports and assists humans in certain circumstances so that rules are created, influenced, controlled, and tailored by a combination of human and machine interaction and intervention. Both kinds of rule-generating approaches have ethical implications.

a) Fully autonomous rule-generating approach

A fully autonomous rule-generating approach would mean the technology produces its own rules and conduct without reference to or intervention from humans. After the initial design and programming by humans, the technology makes its own decisions. This is “machine learning” or “dynamic learning systems” whereby the machine relies on its own databank and experiences to generate future rules and conduct.¹³⁶¹ Fully autonomous weapons systems, for example, would have independent thinking capacity as regards acquiring, tracking, selecting, and attacking human targets in warfare based on previous experience of military scenarios.¹³⁶² Such an approach presents challenges.

There is uncertainty and unpredictability in the rules that a fully autonomous weapons system would generate beyond what it has been designed to do, so that it would not comply with international humanitarian law or ethics. In the civilian sphere, fully autonomous technology may generate rules that adversely impact on human self-worth and progress by causing human redundancies, unemployment, and income instability and inequality. Adverse impact on human self-worth and progress, and uncertainty and unpredictability in the rule-generating process are contrary to what is fundamentally beneficial to

¹³⁶¹ See, P M Asaro, ‘Roberto Cordeschi on Cybernetics and Autonomous Weapons: Reflections and Responses’ (2015) 3 *Paradigmi. Rivistadi critica filosofina* 83-107, 96-98; M J Embrechts, F Rossi, F-M Schleif, and J A Lee, ‘Advances in artificial neural networks, machine learning, and computational intelligence’ (2014) 141 *Neurocomputing* 1-2.

¹³⁶² See, *Report of the ICRC Expert Meeting, Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects* (9 May 2014) (‘2014 ICRC Report’); *Report of the ICRC Expert Meeting, Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons* (15-16 March 2016) (‘2016 ICRC Report’); O Ulgen, ‘Autonomous UAV and Removal of Human Central Thinking Activities: Implications for Legitimate Targeting, Proportionality, and Unnecessary Suffering’ (forthcoming) 1-45.

humankind; such a process cannot produce rules that are inherently desirable, doable, valuable, and capable of universalisation. A perverse “machine subjectivity” or “machine free will” would exist without any constraints.

b) Human-machine rule-generating approach

A human-machine rule-generating approach currently exists in both the civilian and military spheres. IBM, for example, prefers the term “augmented intelligence” rather than artificial intelligence because this better reflects their aim to build systems that enhance and scale human expertise and skills rather than replace them.¹³⁶³ The technology is focused on practical applications that assist people in performing well-defined tasks (e.g. robots that clean houses; robots working with humans in production chains; warehouse robots that take care of the tasks of an entire warehouse; companion robots that entertain, talk, and help elderly people maintain contact with friends, relatives, and doctors). In the military sphere, remotely controlled and semi-autonomous weapons combine human action with weapons technology. Human intervention is necessary to determine when it is appropriate to carry out an attack command or to activate an abort mechanism.

This kind of rule-generating approach keeps the human at the centre of decision-making. But what happens if there are interface problems between the human and machine (e.g. errors; performance failures; breakdown of communication; loss of communication link; mis-coordination)?¹³⁶⁴ This may prove fatal in human-weapon integrated systems reliant on communication and co-ordination, and a back-up system would need to be in place to suspend or abort operations. What happens if the technology is hacked to produce alternative or random rules that cause malfunction, non-performance, or harmful effects? The same problem applies to fully autonomous technology and seems a good reason for restricting use and performance capability to set tasks, controlled scenarios or environments where any potential harm is containable.

Difference between “human will” and “machine will”

Kant defined autonomy of will as “the property the will has of being a law to itself (independently of every property belonging to the object of volition)”.¹³⁶⁵ This may sound chaotic and advocating freedom for humans to do as they please

¹³⁶³ F Rossi, ‘Artificial Intelligence: Potential Benefits and Ethical Considerations’, Briefing Paper to the European Union Parliament Policy Department C: Citizens’ Rights and Constitutional Affairs European Parliament (October 2016)
<[http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI\(2016\)571380_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_EN.pdf)>
accessed 26 August 2017.

¹³⁶⁴ P M Asaro, ‘Roberto Cordeschi on Cybernetics and Autonomous Weapons: Reflections and Responses’ (2015) 3 *Paradigmi. Rivistadi critica filosofina* 90-91.

¹³⁶⁵ I Kant, *The Moral Law: Kant’s Groundwork of the Metaphysics of Morals* (H.J. Paton tr, Hutchinson & Co 1969) 101 [440].

but it is the starting point to explaining how morals come about and how humans should conduct themselves. The ultimate aim of morality is freedom and, therefore, whether conduct is right or wrong is dependent on the extent to which it achieves freedom. If doing something enhances our freedom and can also be universalised to enhance the freedom of others, then it becomes a moral action.

Kant's autonomy of will is hard to transpose into technology because it is reliant on concepts such as self-worth, dignity, freedom, and interaction. A machine would not have a sense of these concepts or be able to attach value to them. "Human will" develops through character and experience to inform moral conduct. "Machine learning" or "dynamic learning systems" that generate rules and conduct based on a databank of previous experiences may resemble a form of "machine will" that makes ethical choices based on internally learned rules of behaviour.¹³⁶⁶ But the human will is much more dynamic, elusive, and able to cope with spontaneity in reaction to novel situations which sit outside rule-based behavioural action and derive from human experience and intuition.

Autonomy of will requires inner and outer development of the person to reach a state of moral standing and be able to engage in moral conduct. This is suggestive of an innate sense of right and wrong. The inner aspect requires adoption and adherence to principles that enhance self-worth and dignity in our person without falling to temptation, personal desires, or external coercion. Examples include avoiding immoral conduct, constantly striving to move from a state of nature to an improved rightful or lawful condition.¹³⁶⁷ By enhancing our self-worth and dignity these principles enable us to function freely as rational beings with autonomy of will. The outer aspect is controlled by principles that enable interaction with others and are capable of universalisation. For example, we accept and abide by the general principle that human interaction should be conducted without resorting to violence. In adhering to this principle we are not just motivated by self-preservation but also a higher norm of preserving freedom; if we start conducting our affairs through violence our interaction will become unstable, unpredictable, and unable to guarantee personal freedom or that of others. Can machines emulate this sort of autonomy?

Artificial intelligence in autonomous weapons may allow machine logic to develop over time to identify correct and incorrect action, showing a limited sense of autonomy. But the machine does not possess a "will" of its own nor does it understand what freedom is and how to go about attaining it by adopting principles that will develop inner and outer autonomy of will. It has no self-determining capacity that can make choices between varying degrees of right

¹³⁶⁶ M O Riedl, 'Computational Narrative Intelligence: A Human-Centered Goal for Artificial Intelligence' (2016) CHI'16 Workshop on Human-Centered Machine Learning, May 8, 2016, San Jose, California, USA; M O Riedl and B Harrison, 'Using Stories to Teach Human Values to Artificial Agents' (2015) Association for the Advancement of Artificial Intelligence.

¹³⁶⁷ I Kant, *The Metaphysics of Morals* (Mary Gregor tr and ed, CUP 1996) 173-218.

and wrong. The human can decide to question or go against the rules but the machine cannot, except in circumstances of malfunction and mis-programming. It has no conception of freedom and how this could be enhanced for itself as well as humans. The machine will not be burdened by moral dilemmas so the deliberative and reflective part of decision-making (vital for understanding consequences of actions and ensuring proportionate responses) is completely absent.

There is a limited sense in which artificial intelligence and robotics may mimic the outer aspect of Kant's autonomy of will. Robots may have a common code of interaction to promote cooperation and avoid conflict among themselves. Autonomous weapons operating in swarms may develop principles that govern how they interact and coordinate action to avoid collision and errors. But these are examples of functional, machine-to-machine interaction that do not extend to human interaction, and so do not represent a form of autonomy of will that is capable of universalisation.

Trust and the technology

When we talk about trust in the context of using artificial intelligence and robotics what we actually mean is reliability. Trust relates to claims and actions people make and is not an abstract thing.¹³⁶⁸ Machines without autonomy of will, in the Kantian sense, and without an ability to make claims cannot be attributed with trust. Algorithms cannot determine whether something is trustworthy or not. So trust is used metaphorically to denote functional reliability; that the machine performs tasks for the set purpose without error or minimal error that is acceptable.

But there is also an extension of this notion of trust connected to human agency in the development and uses to which artificial intelligence and robotics are put. Can we trust the humans involved in developing such technologies that they will do so with ethical considerations in mind (i.e. limiting unnecessary suffering and harm to humans, not violating fundamental human rights)? Once the technology is developed, can we trust those who will make use of it to do so for benevolent rather than malevolent purposes? These questions often surface in debates on data protection and the right to privacy in relation to personal data trawling activities of technologies. Again, this goes back to what values will be installed that reflect ethical conduct and allow the technology to distinguish right from wrong.

¹³⁶⁸ O'Neill, *Autonomy and Trust in Bioethics* (CUP 2002).

The difference between machine logic and human judgment

When we compare machines to humans there is a clear difference between the logic of a calculating machine and the wisdom of human judgment.¹³⁶⁹ Machines perform cost effective and speedy peripheral processing activities based on quantitative analysis, repetitive actions, and sorting data (e.g. mine clearance; and detection of improvised explosive devices). They are good at automatic reasoning and can outperform humans in such activities. But they lack the deliberative and sentient aspects of human reasoning necessary in human scenarios where artificial intelligence may be used. They do not possess complex cognitive ability to appraise a given situation, exercise judgment, and refrain from taking action or limit harm. Unlike humans who can pull back at the last minute or choose a workable alternative, robots have no instinctive or intuitive ability to do the same. For example, during warfare the use of discretion is important to implementing rules on preventing unnecessary suffering, taking precautionary measures, and assessing proportionality. Such discretion is absent in robots.¹³⁷⁰

How will artificial intelligence and robotics engage in moral reasoning in order to act ethically? Should the technology possess universal or particular moral reasoning? Ongoing developments in the civilian and military spheres highlight moral dilemmas and the importance of human moral reasoning to mediate between competing societal interests and values. Companion robots may need to be mindful of privacy and security issues (e.g. protection and disclosure of personal data; strangers who may pose a threat to the person's property, physical and mental well-being) related to assisting their human companion and interacting with third parties (e.g. hospitals; banks; public authorities). Companion robots may need to be designed so that they do not have complete control over their human companion's life which undermines human dignity, autonomy, and privacy. Robots in general may need to lack the ability to deceive and manipulate humans so that human rational thinking and free will remain. Then there is the issue of whether fully autonomous weapons should be developed to replace human combatants in the lethal force decision-making process to kill another human being. Is there a universal moral reasoning that the technology could possess to solve such dilemmas? Or would it have to possess a particular moral reasoning, specific to the technology or scenario?

6 September 2017

¹³⁶⁹ J Weizenbaum, *Computer Power and Human reason: from judgment to calculation* (1976).

¹³⁷⁰ See, E Lieblich and E Benvenisti, 'The Obligation to Exercise Discretion in Warfare: Why Autonomous Weapons Systems Are Unlawful', in *Autonomous Weapons Systems Law, Ethics, Policy* (N. Bhuta, S. Beck, R. Geiß, Liu Hin-Yan, C. Kreß eds., 2016).

University College London (UCL) – Written evidence (AIC0135)

- **A. Summary**

-
- *Current state of AI and its development*
-
- 1. While AI brings challenges and risks, it is largely a force for good and has a wide variety of applications, with the potential, for example, to deliver significant benefits in healthcare. **It is in the public interest to foster the opportunities provided by AI** and address its drawbacks to maximise the potential public good. This will require collaboration between Government, academia, industry and other stakeholders.
- 2. **Further cross-disciplinary and large-scale AI research is needed**, including to develop ethical AI technologies and to build an evidence base for how individuals understand and work with AI systems.
 - UKRI should support AI research through a) cross-disciplinary grants or setting up research centres, b) providing funding to support the creation of large-scale datasets for research and c) facilitating the availability of health-related datasets (in a suitable form) for research.
 -
- 3. The **development of appropriate technical skills** will be key to harnessing the potential of AI, while keeping key ethical and societal considerations in mind.
 - The UK's industrial strategy should include strategies to invest in a pipeline of individuals with computer science skills, and upskill the workforce, including by a) increasing mobility and collaboration between academia and industry, b) replicating successful partnerships between training providers, sector bodies, business associations and employers, and c) increasing Government's technical expertise.
 -
-
- *Social impacts of AI*
-
- 4. Applications of AI can impact the public either visibly, through the automation of processes, or invisibly, subtly shaping a person's environment without their knowledge.
 - There is a duty to **educate the public in the benefits and challenges of AI**, including through education programmes.
 -
-
- 5. It is important to ensure that **individuals' rights are protected** when decisions are made using their data.

- The Government should amend the automated decision safeguards in the UK Data Protection Bill to protect individuals against unfair AI systems.
 - The Government should ensure that there are regulatory frameworks giving people the ability to both understand and input into the process whereby their data are used to create a personalised profile.
-
6. Given the potential for AI to be used in non-transparent, unfair or biased ways, **ethical approaches need to be at the heart of AI.**
 - Ethical standards for AI systems should be established by an independent, interdisciplinary body.
 - Public procurement requirements specifying ethical standards for AI systems should be introduced.
 - The Government should commission research leading to the development of Equality Impact Assessments for the procurement and management of AI systems.
 - Privacy Enhancing Technologies should be put in place to ensure that AI systems respect privacy, and awareness of these technologies' effectiveness should be raised.

B. Introduction

1. UCL is pleased to make a submission to the House of Lords Artificial Intelligence (AI) Select Committee in response to its call for evidence on the implications of AI.
2. UCL is London's leading multi-faculty university, with more than 11,000 staff and 38,000 students from 150 different countries. It is one of the leading academic data science centres in the UK, with a number of departments engaging in data science and AI research. UCL is also one of the five founding partners of the national Alan Turing Institute for Data Science. Current research areas at UCL include the use of machine learning to identify patterns in data and process natural language; the development of techniques to design privacy into data processing systems; and the responsible use of machine learning to support decisions by public sector organisations. UCL is also involved in a number of events relating to AI, such as the Data for Policy conference¹³⁷¹ this September and a scientific meeting¹³⁷² at the Royal Society in October on the growing ubiquity of algorithms in society.

¹³⁷¹ <http://dataforpolicy.org/>

¹³⁷² <https://royalsociety.org/science-events-and-lectures/2017/10/algorithms-society/>

3. This document is based on contributions from the following members of the UCL community: George Danezis, Emiliano De Cristofaro, Zeynep Engin, Sebastian Riedel, Pontus Stenetorp, Raphael Toledo, Philip Treleaven, Johannes Welbl (Department of Computer Science); Miguel Rodrigues (Department of Electronic and Electrical Engineering); Sylvie Delacroix, Richard Moorhead (UCL Laws); Geraint Rees (Faculty of Life Sciences); Michael Veale (Department of Science, Technology, Engineering and Public Policy (STEaPP)); Sofia Olhede, Patrick Wolfe (Department of Statistical Science, UCL Big Data Institute).

C. The current state of AI and its development

4. AI provides computers with the ability to learn and make decisions without explicit programming. There have been remarkable developments in recent years on various AI tasks, including computer vision, speech recognition, translation, image and object recognition, recommender systems and games.
5. It is widely accepted that the progress in AI has been mainly driven by 1) improved algorithms; 2) the availability of vast amounts of data; 3) the availability of massive computing power. In addition the rapid increase in systems using data as part of the internet of things (network of intercommunicating objects connected to the internet) has strongly accelerated AI's development. Combined, these factors have led to systems outperforming humans at a number of specific tasks, which they were unable to do just a few years ago.¹³⁷³
6. Building on these advances, we expect the AI field to develop substantially in the coming years and cause disruption across all sectors. However, it is unclear whether there will be much progress from the current narrow AI systems (which can work on a specific task) to general AI ones (able to apply intelligence to various problems). The current understanding of narrow AI systems is poor, and the ability to understand the fundamentals of these algorithms may accelerate the development of AI systems. To advance the understanding and development of AI systems, there is a need to create **research environments that support researchers across disciplines to investigate AI applications in data-heavy sectors of public interest**, such as healthcare. **UKRI has a key role in delivering cross-disciplinary grants or setting up multi-disciplinary research centres**; the involvement of the corporate sector in such centres should also be encouraged.

¹³⁷³ <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

7. Technology is making AI more accessible to non-experts. Major IT companies such as Facebook and Google are likely to continue driving progress through ready-to-use AI software (such as TensorFlow and PyTorch). Other emerging technologies, such as Blockchain and the Internet of Things, are providing the infrastructure for more user-friendly, automated and smart data collection, and secure storage and processing of sensitive data. For example, the EPSRC-funded UCL Urban Dynamics Lab is creating an online community platform to make data, analytics and expertise more accessible to all groups. The aim is to create a secure and trusted public infrastructure for data from a range of sources and link it to the platform, to allow researchers, policy makers and non-expert public users to carry out advanced data analysis, including using AI. Initiatives such as these will be integral to facilitating the use of AI by non-experts, extending its impact across society while maintaining data security.

D. The impact of AI on society

8. The impact of AI technologies on society already is and will continue to be considerable. Applications of AI can impact the public either visibly, through the automation of processes, or invisibly, subtly shaping a person's environment without their knowledge, such as AI-based personal profiles determining insurance tariffs, job offers or personalised advertisement. AI-driven applications can learn to mimic human judgements, echoing human stereotypes and biases. If widely deployed, AI algorithms have the potential to define new norms or correct existing ones so it will be crucial to ensure that ethical frameworks are in place to protect individuals and their data (see section J).
9. The ability to automate semi-intelligent tasks will change the production model in many countries. AI is expected to have an effect on the workforce,¹³⁷⁴ possibly removing a large percentage of jobs,^{1375,1376,1377} (numbers as varied as 30% to nearly 50% are quoted for the industrialised world). Automation increases the risk of unemployment, particularly for low-skilled workers, and potentially high-skilled workers in the future. While employers can benefit considerably from automation, employees, whose jobs may be at risk, will benefit the least.
10. In order to harness the potential of AI and improve the capability of the workforce, **the UK's industrial strategy should include strategies to**

¹³⁷⁴ <http://www.pwc.co.uk/services/economics-policy/insights/the-impact-of-artificial-intelligence-on-the-uk-economy.html>

¹³⁷⁵ http://pwc.blogs.com/press_room/2017/03/up-to-30-of-existing-uk-jobs-could-be-impacted-by-automation-by-early-2030s-but-this-should-be-offse.html

¹³⁷⁶ <https://www.wired.com/brandlab/2015/04/rise-machines-future-lots-robots-jobs-humans/>

¹³⁷⁷ http://www.oxfordmartin.ox.ac.uk/downloads/reports/Citi_GPS_Technology_Work_2.pdf

upskill the workforce and invest in a pipeline of talent:

- a. Expertise should be shared and improved by **increasing mobility and collaboration between academia and industry**, and by **replicating successful partnerships** between training providers, sector bodies, business associations and employers.
- b. There is a need to increase the **technical expertise of Government in translating codes of practice into law and interpreting existing legal frameworks in a new technical context**. AI will need to be regulated according to the level of risks in specific contexts, which will vary, and technical expertise is required to ensure this.
- c. Given the high demand for those with computer science skills, it will be important to **ensure that the pipeline of such individuals, for both academia and industry, is well-populated**. Universities will play a key role in exposing those educated in the sciences and engineering to an AI curriculum that they can draw upon during their career. This would enable the UK to capture a greater market share of AI creators.

E. Preparing the public for more widespread use of AI

11. Individuals' rights need to be protected to prevent unfair use of their data. Under the General Data Protection Regulation, an individual has certain rights when decisions are made using their data, if the decisions are *fully automated* and *significant*.¹³⁷⁸ However, few highly significant decisions are fully automated – often they are used as decision support, for example in detecting child abuse or assessing candidates for a job, yet AI systems can still bias these decisions. For example, a CV filtering system that uses past success rates for job applicants is likely to replicate any biases that existed when those applicants were assessed manually in the past. Additionally, few fully automated decisions are individually significant, even though they might be over time; for example one advert may not be significant, but someone's environment can be shaped significantly by adverts over time. Given the potential for unfair data use in cases such as these, **the Government should amend the automated decision safeguards in the UK Data Protection Bill to explicitly protect individuals against unfair and in transparent AI systems they face in their day-to-day lives.**

F. The public's understanding of AI

¹³⁷⁸ Lilian Edwards and Michael Veale (forthcoming). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. Duke Law and Technology Review. Available on SSRN <http://ssrn.com/abstract=2972855>

12. There is a duty to educate the public in the benefits and challenges of AI, as these systems increasingly permeate our lives. The Royal Society has explored this issue,¹³⁷⁹ finding considerable public fear around AI, especially of loss of jobs and a restriction of opportunities for individuals. If a number of AI examples developed badly, there could be considerable public backlash, as happened with genetically modified organisms.
- a. **UKRI should support research into how individuals understand and work with AI systems, through cross-disciplinary grants**, to improve understanding of how the public currently relates to AI, and inform efforts to raise awareness.
 - b. **Education programmes and activities, involving Government, academia, industry and the media, should be used to improve the public's understanding of AI.** It will be important to raise awareness of the AI applications that are already improving people's lives (such as search engines), and the profiling mechanisms used for selecting personalised content. Engaging students will be important, and lessons may be learned from the Royal Society's work on how to harness expertise from businesses and academia to support computing education in schools.³
 - c. **It should be required that users can easily expose and understand systems affecting them, for example why a particular advert was shown to them**, in order to reduce any undue public fear around AI.
13. It will be important to engage the public with decision-making and trade-offs in AI. In many cases there are practical ways to explain what AI does,⁸ and more that might be possible with proper research (for example, see DARPA's recent Explainable Artificial Intelligence research programme)¹³⁸⁰. However many AI methods are complex, which can make the reasons for decision-making unclear. Indeed, AI methods that are more complex and less transparent often perform better. In such cases, it will be important to engage the public to decide what price is worth paying for transparency.

G. Sectors most likely to benefit from AI

14. Any sector that generates or has access to large amounts of data will benefit from AI. The main AI players in the private sector are thus mostly web companies with a large user base, who can collect data automatically at very little cost. Currently, this mostly benefits the advertisement industry, but other data-intensive industries or research fields, such as public administration, healthcare or the biosciences, can benefit as well. As AI technology develops, its use is likely to expand across a number of

¹³⁷⁹ <https://royalsociety.org/topics-policy/projects/machine-learning/>

¹³⁸⁰ <https://www.darpa.mil/program/explainable-artificial-intelligence>

industries.

15. There is great potential for AI in healthcare, by making services more efficient by anticipating demand¹³⁸¹ and to support diagnosis and clinical decision making. For example, work at the UCL Centre for Health Informatics & Multiprofessional Education (CHIME) has explored the use of AI in decision support systems for the diagnosis and treatment of breast cancer. This system has been trialled at the Royal Free Hospital Breast Unit and showed excellent agreement between the recommendations by the AI system and decisions made by a multidisciplinary team. This illustrates the great potential for AI to improve health. Collaborations between universities, the NHS and industry, such as Google DeepMind (which has seen a lot of press interest¹³⁸²), also have tremendous potential to improve health by improving diagnosis and treatment, but care needs to be taken to ensure the privacy of individuals' data.

1)

16. There is a need to develop new ways to evaluate the effectiveness of AI technologies in healthcare, keeping in mind that such technologies form part of a clinical care pathway. For example, UCL is in collaboration with Google DeepMind and the Royal Free Hospital to evaluate the impact of an app used to alert clinicians to potential cases of acute kidney injury¹³⁸³. It is important to promote the development of skills to support such evaluation.

17. Since the Government collects and processes citizen data, it can benefit greatly from AI, particularly in the involvement of the public in decision making; providing 'intelligent' service delivery; and increasing efficiency of the design and operation of the public infrastructure. For instance, if personal data from front-end services (health records, education, crime history, and so on) can be linked in a secure and trusted way to other sources (such as data on banking transactions or transport), AI systems can provide the civil service with deeper insights for decision making. This has the potential to support a wide range of decisions, on areas ranging from potential child abuse to local planning.¹³⁸⁴ In this way AI systems can provide a great deal of supporting information to the civil service, lowering the burden of work and making Government processes more efficient.

H. The data-based monopolies of large corporations

¹³⁸¹ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf

¹³⁸² <http://www.wired.co.uk/article/deepmind-nhs-streams-deal>

¹³⁸³ <https://f1000research.com/articles/6-1033/v1>

¹³⁸⁴ <http://www.nesta.org.uk/publications/wise-council-insights-cutting-edge-data-driven-local-government>

18. Given the risk of large corporations holding data-based monopolies, there is a need to support SMEs, researchers and the public to access data and use AI systems. There is a risk of a data disparity in research, where private companies have access to more funding and data sources so carry out research that would be impossible in the public sector. There is also value in public sector data (such as health-related data) being shared in certain circumstances.

- a. **UKRI has a role to play in providing funding to support the creation of large-scale datasets for research, and in facilitating the availability of health-related datasets for research**, in a suitable form that respects privacy. UK Biobank provides a good example of the value of making health-related data available for research.
- b. **The Government should explore models for data repositories and regulatory incentives for storing de-sensitised and de-identified private sector data.** Such a repository could be modelled on the Administrative Data Research Network (ADRN) and the UK Data Archive.
- c. **The Government should provide a secure infrastructure for holding key datasets for AI systems and making them as widely available as necessary or appropriate.** This infrastructure should include expertise on de-identification and secure access. It might include data on online and mobile behaviour, language, robotics and automated vehicles. There could also be a parallel, separate infrastructure for NHS data.
- d. **The Government should continue to strike a better balance between copyright and the public interest by supporting access to data for research.** The UK is already ahead of the competition in this regard, with its copyright exception for data mining in force since 2014 (which allows researchers to make copies of copyright material for use in data analysis for non-commercial research)¹³⁸⁵.
- e. **The Government should continue to pursue and enable open data initiatives.** For example, data.gov.uk provides free access to data from Government departments and public bodies, and the app Citymapper uses open data from Transport for London and other sources to facilitate the population's use of public transport.

I. Managing and safeguarding data for the public good and a well-functioning economy

19. Individuals' privacy should be safeguarded in AI systems.

¹³⁸⁵ <https://www.jisc.ac.uk/guides/text-and-data-mining-copyright-exception>

- a. **Privacy Enhancing Technologies (PETs) should be put in place to ensure that AI systems respect privacy.** There are a number of PETs proven to be able to process statistical data without revealing it,^{1386,1387,1388} to protect privacy even in highly sensitive settings.¹³⁸⁹ UCL has proposed a number of techniques to efficiently get aggregate statistics over encrypted data, and this is a thriving research field.^{17,1390} In cases where such techniques are not applicable, further research will be needed to develop suitable methods rather than compromise privacy.
- b. **It is important to raise the public's and decision makers' awareness of PETs,** to gain public trust for the ethical use of AI (see section F).

20. Initiatives to improve public trust in the use of data have value. There is the possibility of individuals collating their data and forming a mutual organisation to manage a 'data trust', with conditions for how the data can be used.¹³⁹¹ The aim of such a trust is to give individuals, such as patients, more rights over their data, rather than the model whereby patients' data can be shared without their awareness, based on their implicit consent. Models incorporating elements of data trusts may be useful to gain public trust in the ethical use of AI.

J. The ethical implications of AI

21. While AI has pitfalls, it is largely a force for good: it can be a key technology when integrated with a wide array of applications, providing numerous possibilities to improve the quality of human life. For example there is tremendous potential in its use in healthcare (see section G). It is in the public interest to foster the opportunity provided by AI, and address its challenges.

¹³⁸⁶ Apostolos Pyrgelis, Emiliano De Cristofaro, Gordon J. Ross: Privacy-Friendly Mobility Analytics using Aggregate Location Data, 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2016) <https://arxiv.org/pdf/1609.06582.pdf>

¹³⁸⁷ Luca Melis, George Danezis, Emiliano De Cristofaro: Efficient Private Statistics with Succinct Sketches
23rd Network and Distributed System Security Symposium (NDSS 2016)
<http://arxiv.org/pdf/1508.06110.pdf>

¹³⁸⁸ Julien Freudiger, Emiliano De Cristofaro, Alex Brito: Controlled Data Sharing for Collaborative Predictive Blacklisting, 12th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA 2015) <https://arxiv.org/pdf/1502.05337.pdf>

¹³⁸⁹ <https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design>

¹³⁹⁰ Tariq Elahi, George Danezis, Ian Goldberg: PrivEx: Private Collection of Traffic Statistics for Anonymous Communication Networks. ACM Conference on Computer and Communications Security 2014: 1068-1079

¹³⁹¹ <https://www.theguardian.com/media-network/2016/jun/03/data-trusts-privacy-fears-feudalism-democracy>

22. It is important to consider how values are used to design algorithms used to make decisions. **It is worth investigating the potential of Interactive Machine Learning, which allows end-users to retain an active role in the learning process of the AI technology, allowing for their changing values to be taken into account.**
23. Highly personalised profiling gives rise to significant risks that users could be manipulated into certain preferences. The Government should ensure that there are **regulatory frameworks giving people the ability to both understand and input into the process whereby their data are used to create a personalised profile.**
24. The Government should use public procurement, research funding and industrial strategy to foster the UK's markets and skills in fair, transparent and ethical AI:
- a. **Public procurement requirements should specify ethical standards for AI systems in areas such as transport, health, policing and security.**
 - b. **UKRI and other research funders should support the development of ethical AI technologies.**
 - c. **Ethical standards for AI systems should be established by an independent, interdisciplinary body**, such as the data use stewardship body laid out by the recent Royal Society and British Academy report on 'Data Management and Use'¹³⁹². For example, modern AI systems are largely probabilistic and work in most cases but not always, so scenarios where failure would be catastrophic should be avoided. In addition, AI-based decisions affecting an individual should be transparent. There are some provisions for this in the General Data Protection Regulation, but these provisions might need several loopholes to be closed to be effective (see section E).⁸
 - d. **Investments should be made in improving transparency** in AI; **improving reliability** in AI to support its use in challenging, real world environments; and **improving fairness and eliminating biases** in AI (for example AI systems used in the understanding of language have been shown to exhibit gender bias).
25. AI is increasingly used in the public sector for purposes including policing, taxation, justice, child protection and emergency response; ensuring fairness in these cases is crucial.¹³⁹³ **The Government should**

¹³⁹² <https://royalsociety.org/topics-policy/projects/data-governance/>

¹³⁹³ Michael Veale (2017). Logics and practices of transparency in real-world applications of public sector machine learning. 2nd Workshop in Fairness, Accountability and Transparency in Machine Learning (FAT/ML 2017). Available from <http://arxiv.org/abs/1706.09249>

commission research leading to the construction of Equality Impact Assessments for the procurement and management of AI systems, as the current framework is insufficient.

26. Efforts should be made to ensure the responsible use of machine learning in the public sector. Machine learning situations have been developed in low-stakes private sector environments, like online shopping. Public sector applications differ strongly, and tend to focus on modelling rare, high stakes events, such as child abuse, burglaries, tax or benefit fraud, and so on. Given this, **skills within Government should be developed to ensure that AI is used responsibly in the public sector.**
27. The level of transparency required depends on the purpose the AI system is being used for. Black boxing (the use of AI systems that do not easily allow an explanation to be found for why a result has been obtained) may be acceptable for low-level tasks but not for higher level ones. For example, the use of AI systems to support decisions that may affect an individual (such as medical diagnosis) should entail some degree of 'explanation' behind the decision. This ensures the ability to verify the basis behind important decisions, and also learn from it.

K. The role of the Government

28. The Government should play a role in educating the public, protecting citizens' rights, supporting AI research and access to data, fostering competition between companies, and amending legislation (e.g. the Data Protection Bill) where necessary or appropriate to support these endeavours. For specific recommendations to Government, see paras 10-12, 18, 23-26.

L. The work of other countries or international organisations

29. Internationally there is a great deal of investment in digital Government services, including by most developed countries. Examples include Singapore's Sigpass single signon system providing access to a holistic range of Government services; the UK's 'digital by default' strategy; and Germany's 'Bundesagentur für Arbeit' virtual labour market platform to reintegrate jobseekers into the labour market. Additionally the European Commission has an ISA2 Programme,¹³⁹⁴ which supports the development of digital solutions that enable public administrations, businesses and citizens to benefit from interoperable cross-border and cross-sector public services. There is value in learning from the successes (and any negative implications) arising from such initiatives.

¹³⁹⁴ https://ec.europa.eu/isa2/isa2_en

6 September 2017

**Sheena Urwin and Marion Oswald – Written evidence
(AIC0068)**

[Submission to be found under Marion Oswald](#)

Michael Veale – Written evidence (AIC0065)

AI Committee Submission

Michael Veale, researcher in responsible public sector machine learning, UCL STEaPP

University College London, Department of Science, Technology, Engineering and Public Policy // m.veale@ucl.ac.uk

The government should use public procurement, research funding and industrial strategy to foster the UK's markets and skills in fair, transparent, and ethical AI

This should be done by:

- Using public procurement requirements to specify ethical standards for AI systems in areas such as transport, health, policing and security.
- Using UKRI and similar bodies to support the development of relevant ethical AI technologies.

These standards should be established by an independent, interdisciplinary body, such as the data use stewardship body laid out in the recent *Royal Society and British Academy* report on *Data Management and Use: Governance in the 21st Century*. There is political will for such body: the 2017 Conservative Party Manifesto specified a *Data Use and Ethics Commission*, and a similar body was recommended by two previous Select Committees (The Big Data Dilemma; Robotics and Artificial Intelligence).

In particular, investments will need to be made in

- **Transparency and AI:** AI systems are often described as a 'black box', but in many cases this need not be so (Edwards and Veale, forthcoming). There are many practical ways to usefully explain what AI does that already exist, and more that might be possible with proper research (for example, see DARPA's recent *Explainable Artificial Intelligence* research programme).
- **Fairness and AI:** AI systems are primarily very data heavy, and such data is not neutral. Indeed, much data encode problematic biases. Especially when being asked to do a complex task, biases can creep in. For example, in the understanding of language, the most adept AI systems today have been shown to exhibit gender bias. When understanding analogies, for example, 'man' is to 'woman' as 'computer programmer' is to 'homemaker' — one of many biases that it is unlikely to be desirable to reproduce.
- **Reliability and AI:** AI is being touted for use in highly consequential systems. It is naturally desirable that these systems work, particularly in safety critical environments, but it is far from guaranteed, particularly when

these environments change. AI that can work in challenging, real world environments will be highly sought after in practice.

The use of AI in the public sector requires the development of AI-specific Equality Impact Assessments to meet the Public Sector Equality Duty in the Equality Act 2010

AI, including statistical systems powered by machine learning, suffers from issues of bias and fairness that requires careful and methodological social and technical investigation. AI is being increasingly used in the public sector for a range of purposes, including policing, taxation, justice, child protection, and emergency response (Veale, 2017). The current framework of examining and documenting consequential practices using Equality Impact Assessments in the public sector is insufficient to examine these practices. The Government should commission research leading to the construction of Equality Impact Assessments fit for the procurement and management of AI systems.

The automated decision safeguards in the Data Protection Bill should be amended to explicitly protect individuals against unfair and intransparent AI systems they may face in their day-to-day lives.

As they exist at the moment, the automated decision safeguards place too much emphasis on the requirement for a decision to be *fully automated* and *significant* before they are applicable (Edwards and Veale, forthcoming). The problem here is that few highly significant decisions are fully automated — often, they are used as decision support, for example in detecting child abuse. Additionally few fully automated decisions are individually significant, even though they might be over time. Advertising might not be significant from the perspective of delivery of a single advert, but how someone's environment is shaped by adverts over time is likely very significant. Furthermore, 'automated' has in the past, by other courts, been read very strictly (Wachter et al. 2017), and this seems inapplicable to a world where humans and machines are interacting constantly.

There is a need here for a few different provisions:

- Explanation of systems where AI is only one part of the final decision. This is, for example, a provision in the recent Digital Republic Act in France.
- Systems that do not require the use of a 'right' to have them explained, but instead are incentivised, through regulatory or other means, to be *transparent by design*. In essence, this involves different ways of ensuring systems' transparency throughout their functioning so that individuals do not feel the need to seek the redress of an individual right. Transparency by design should be proportional — not all systems can, or need to be, transparent, but justification should be given for the level of transparency provided and the trade-offs made in relation to the state-of-the-art.

The Government should commission research into how individuals understand and work with AI systems.

The understanding of AI systems is not a solely technical challenge. Individuals and groups understand these technologies in very different ways: for example, over-relying or under-relying on their outputs. The Government should ensure that research funding for AI supports deep collaboration with the social sciences and humanities.

In the face of the risk of data-based monopolies, the Government should explore repository models and regulatory incentives for storing de-sensitised and de-identified private sector data for SMEs, researchers and the public to train AI systems. Such as repository could be modelled after the ADRN and the UK Data Archive.

Firms accumulating a great deal of data have disproportionate market power in building AI adept at particular tasks. Some laws already provide some frameworks for breaking these data-based monopolies in respect to *personal data*, which might include location data, voice data, or data from household devices, such as the GDPR's *right to data portability*. Recent draft guidance (from the Article 29 Working Party) limits the effectiveness of these rights over *inferred data*, which is potentially highly limiting of the ability of these rights to influence AI systems, and it may be necessary to consider whether this right is sufficient.

As an issue of competition (the French and the German competition authorities released a joint report on data and competition law in 2016; the UK's Competition and Markets Authority released a document on Consumer Data in 2015) companies might need solutions to ensure they can grow whilst being compliant.

The Government should provide a secure infrastructure for holding and accumulating key datasets for AI systems and making them as widely available as necessary or appropriate. Such an infrastructure should include expertise on deidentification and secure access, such as the microdata access regimes already present in statistical agencies around the world.

This data might include

- Audiovisual data from automated vehicles
- Language and comment data
- Robotics and physical sensor data
- Parallel (but separate) NHS infrastructure for healthcare data
- Online and mobile behaviour data

References

Edwards L and Veale M (forthcoming). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law and Technology Review*. Available on SSRN <http://ssrn.com/abstract=2972855>

Wachter S, Mittelstadt B, Floridi L (2017). Why a right to explanation does not exist in the General Data Protection Regulation. *International Data Privacy Law*.

Veale M (2017). Logics and practices of transparency in real-world applications of public sector machine learning. 4th Workshop in Fairness, Accountability and Transparency in Machine Learning (FAT/ML 2017). Available from <http://arxiv.org/abs/1706.09249>

4 September 2017

Professor Chris Voss – Written evidence (AIC0118)

Impact of Artificial Intelligence in service industries on employee roles

Chris Voss, Professor of Operations Management, Warwick Business School,
Emeritus Professor of Operations Management, London Business School

1. **Background.** In 2016, a multi-country group of researchers from the field of service management, worked to explore the implications for work of the trends in technology, in particular Artificial Intelligence (AI). This submission is based upon our final output which was published this year in the Journal of Business Research.¹³⁹⁵ The focus of our work is mainly on the impact of technology on the employee and the service encounter (service encounter 2.0), but we also consider the impact on the customer.

2. Summary of arguments in the submission

Much writing on AI focuses on the impact on employment; we do not wish to replicate this as current models by organisations such as McKinsey are good. We accept that AI will lead to job losses in many sectors. We also argue that substitution and deskilling of service employees is only one of the possible impacts of such technology. It also has the potential to augment the roles and jobs of service employees and can serve as an enabler of connections and relationships.

The focus of this submission is that an important impact of AI and similar new technologies is that there will be a need for changes in the roles of employees. We argue that it is important to consider the impact of AI and similar technologies on employees from of all technology roles, not just substitution and deskilling. We put forward four roles that are important both for employees and for the effective and positive use of technology; Enabler, Innovator, Coordinator or Differentiator. AI and related technologies also impact customers and we see a parallel set of roles for customers.

For each of these roles positive employee (and customer) experience and performance will be strongly influenced by their *role readiness*.

It is important that we come to understand how, due to the impact of AI and related technologies, the employee role is changing in many service settings. This understanding will be of vital importance to managers and public policy makers to prepare for the future of the human workforce.

¹³⁹⁵ Bart Larivière, David Bowen, Tor W. Andreassen, Werner Kunz, Nancy J. Sirianni, Chris Voss Nancy V. Wunderlich, Arne De Keyser, (2017), "Service Encounter 2.0": An investigation into the roles of technology, employees and customers, Journal of Business Research, 79, 238-246

We now explore this in more detail below.

3. **Technology.** The current debate on technology has been confused by the excessive use of the word “Robot”. In the pure sense of anthropomorphic and manipulative technologies, robots play only a small part in services. The press always putting a picture of a robot in every discussion of AI diverts attention away from the real impact of technologies. We argue that new technology, in particular AI has, three potential roles, all of which imply significant change for employees, organisations and society, but not all have a negative impact on employees. These roles are:

- 3.1. **Augmentation of service employees** – signifies technology’s ability to assist and complement service employees in the service encounter. In popular press, this is often referred to as Intelligence Augmentation (IA), reflecting situations in which technology supports human thinking, analysis and behaviour. In other words, technology can be used in tandem with employees to provide a better service encounter outcome. Technology as augmentation can typically be found with the promise of enhancing employees’ service delivery capacity. Intelligent assistants can help customers find products and can answer simple questions. As a result, employees can spend more time offering specialty knowledge to customers. In a service provider context, healthcare organizations offer one of the most fertile grounds for technology augmentation. Here, Intelligent Assistants are increasingly complementing human care providers. For example, IBM’s Watson now assists medical doctors in diagnosis, whereas service robots are increasingly collaborating with human medical staff in elderly care (van Doorn et al., 2017).

- 3.2. At the same time, advances in AI in robots, sensor fusion, deep learning algorithms and smart devices are causing employees to become obsolete in their traditional service encounter position. Thus, the second role of technology – **substitution of service employees** – reflects the purpose of replacing human (i.e., employee) input in the service encounter. Service employees no longer take active part in the service encounter that becomes fully technology-generated. Technology promises to increase service encounter quality and efficiency, omitting inherent human performance. As intelligent systems are now able to deliver more advanced services, we observe that also higher-level jobs are threatened (Marr, 2016). For example, U.S.-based law firm Baker Hostetler is now making use of an artificially intelligent system, Ross, to help perform legal research and (potentially) replace part of the labour force in the future.

- 3.3. The third role of technology – **network facilitation** – refers to technology acting as an enabler of connections and relationships. Stimulated by the swift development of digital platforms and Internet of

Things, this role is rapidly gaining traction. Clearly, *Network Orchestrators* heavily build on such technologies. Rather than focusing on replacing human employees, these business models seek to use technology as a way to connect multiple entities in the service encounter – both human and technological. These constellations are also referred to as multi-sided markets defined by multiple distinct entities that provide each other, via a platform, with network benefits. Airbnb, for example, uses a technology-based platform to facilitate exchange between private house owners willing to rent their property to travellers. Likewise, Uber's platform connects private drivers and customers in need of transportation. Both Airbnb and Uber do not own physical assets – hotels and cars, respectively – but merely facilitate service exchange through use of network technology.

4. We argue that it is important to consider the impact of AI and similar technologies on employees from of all three technology roles, not just substitution and deskilling. We see four transformed roles for employees in the Service Encounter 2.0 – the employee as an *Enabler*, *Innovator*, *Coordinator* and *Differentiator*. These roles are not mutually exclusive, meaning an employee might take on more than one role. And we recognize that the traditional service employee role – actual delivery of the service – still exists in many services today. The “service employee as the service”-principle will also hold true for some services in the future. Building technological alternatives for every service is not economically viable in all circumstances. For example, some markets/segments might not be technology ready or too narrow to be served by machines/technology. However, it is important that we come to understand how the employee role is already changing in many service settings. This understanding will be of vital importance to managers and public policy makers to prepare for the future of the human workforce.

- 4.1. The first transformed employee role is that of **enabler**. In an enabling function, employees help both customers and technology to perform their respective service encounter roles well. Sometimes customers and/or technology can experience difficulties that lead to negative customer outcomes such as anger, frustration, and dissatisfaction. To prevent this from happening, employees can advise customers beyond the transaction and/or handle conflicts that result from technology failures or customers' incapacity to deal with a certain online interface. Previous research also demonstrated service employees' enabler role to help gain user acceptance of novel technological interfaces. The enabler role is not only relevant for front-line employees in augmentation situations, but back-office workers also have an equally strong enabling role when technology fully substitutes the human front-line.

- 4.2. Employees may act as **innovators** since human capital remains a non- substitutable source of creativity. Actively dealing with customers in augmentation, functioning as the “front-line” for customer contact in substitution and monitoring connections in network facilitation, service employees, directly and indirectly, observe customer behaviours and reactions. This makes employees highly valuable assets in that they can serve as a barometer of the customer environment and actively pinpoint areas for service improvement. Furthermore, machines have shown little creative ability until now. While this is perhaps gradually changing, we posit that employees as part of the service system can still better read customer needs. The important role of employees in innovation is evident in research showing that the more contact employees are involved in the service innovation process, the greater innovation volume and innovation radicalness.
- 4.3. Employees can take on a **coordinator** position in the service encounter. This role becomes increasingly prevalent as complex service systems comprised of multiple actors require active coordination to create successful outcomes. In these situations, employees can function as a leading party to harmonize and manage the interdependencies between different network partners. Also, a single service encounter does not typically stand by itself. Rather, it is often connected to a series of other encounters across multiple channels that together give shape to an overall customer experience. The value of this experience is largely dependent on the consistency and connectedness of each distinct encounter, which can be managed by service employees in a coordinating role.
- 4.4. A final employee role is that of a **differentiator**. The unique position of employees as a means to differentiate is as important as it has always been. Technology is not loyal, and can often be copied easily. Service employees and their skills, however, are less replicable, authentic human touch can help differentiate offerings in the marketplace and display unique brand-building behaviours, customers are people first, and only customer second. Recent research for example, reveals that the need for human touch can be especially relevant in after-sales situations (e.g., service requests and failure handling). It shows that seemingly internet-savvy customers often prefer human contact in after-sales. This illustrates that the optimal balance between “tech” and “touch” must be found for every service encounter situation. In making these decisions, managers should keep in mind that service employees might add a unique dimension to technology, regardless of its functionality.
5. **Transformed Customer roles.** AI and related technologies impact customers as well. Much like employees, customers also take on distinct and

changing roles in the Service Encounter 2.0. These largely mirror those of the employee, and we again distinguish 4 different roles – the customer as an Enabler, Innovator, Coordinator, and Differentiator. These roles are not mutually exclusive and can occur at the same time.

6. Employee Outcomes and Role Readiness

From our above discussion, it is clear that employees (and customers) are now confronted with new roles in the service encounter. These new roles come with significant challenges for both employees and customers. Their ability to perform well (i.e., role performance) and the resulting experiences will largely depend on employee/customer role readiness - a state or condition in which a person is prepared to perform a specific role. This is driven by three factors: role clarity (i.e., does an employee/customer understand what is expected?), ability (i.e., is an employee/customer able to perform as expected?), and motivation (i.e., is an employee/customer willing to perform as expected?). The more an employee is "ready" to excel at one or more of his/her changed roles, and then performs well and feels rewarded for doing so, the more positive employee experience is likely to be. If, on the other hand, an employee is not ready to cope with changed job requirements, this will reflect negatively on role performance and employee experience. Therefore, companies need to invest significantly in preparing employees for their changing role in the service encounter

6.1. Employee role clarity is determined by one's understanding of the expectations that come with a specific service job. Clearly, the above-presented roles of enabler, innovator, coordinator and differentiator set additional job expectations above what is traditionally expected from a service employee. For example, a coordinating role requires employees to manage multiple parties in co-shaping the service encounter process, which is different from traditional dyadic settings. The more an employee is uncertain on how to execute his/her new role and what is expected, the lower job satisfaction and psychological well-being will be. To avoid this negative outcome, managerial socialization processes are important. These allow employees to get familiar with and adopt required behavioural patterns and norms. Clear feedback systems, the development of job guidelines and goal setting are key practices to increase role clarity.

6.2. Employee role ability reflects the extent to which one is able to perform his or her job in line with what is expected. Managerial support and training are key to enhance employee ability. Employees must be equipped with the right skillset to be successful in their new roles. Three abilities are especially relevant in today's service environment: creativity, empathy (i.e., social skills) and digital fluency. Creativity and empathy are two areas where humans are still superior to technology, and are directly linked to the enabler, innovator and differentiator roles. Digital

fluency, which reflects an employee's proficiency and comfort in achieving desired outcomes using technology, is a key qualifier to function in the Service Encounter 2.0. As technology works in combination with human employees, it is important that the latter are able to deal with their novel 'partner'. While important in an enabling role, digital fluency is especially essential in coordinating many of today's (online) service networks. This, however, does not mean that traditional skills needed for service delivery should be neglected in training. In case of a technology breakdown, for example, employees should still be able to step in to guarantee successful service encounter outcomes.

6.3. Employee role motivation reflects an employee's willingness to perform his/her role as expected and is impacted by managerial encouragement processes. The latter entail, for example, enriching job characteristics and the whole of appraisal and reward systems. While decent financial remuneration through basic pay and performance bonuses is essential, performance appraisal, feedback and recognition from customers, colleagues, and managers are equally important motivational triggers. Furthermore, employee empowerment will prove to be an increasingly important motivator – especially when one considers that all of the transformed employee roles require some freedom in dealing with customers and technology.

6 September 2017

Dr Toby Walsh – Written evidence (AIC0078)

Written Submission to
House of Lords
Select Committee on Artificial Intelligence
Prof. Toby Walsh FAA, FAAAI, FEurAI.

1. Pace of technological change.

Recent advances in AI are being driven by four rapid changes: the doubling of processing power every two years (aka Moore's Law), the doubling of data storage also every two years (aka Kryder's Law), significant improvements in AI algorithms especially in the area of Machine Learning, and a doubling of funding into the field also roughly every two years. This has enabled significant progress to be made on a number of aspects of AI, especially in areas like image processing, speech recognition and machine translation. Nevertheless many barriers remain to building machines that match the breadth of human cognitive capabilities. A recent survey I conducted of hundreds of members of the public and as well as experts in the field (<https://arxiv.org/abs/1706.06906>) reveals that experts are significantly more cautious about the challenges remaining.

2. Impact on society.

Education is likely the best tool to prepare the public for the changes that AI will bring to almost every aspect of our lives. An informed society is one that will best be able to make good choices so we all share the benefits. Life-long education will be the key to keeping ahead of the machines as many jobs start to be displaced by automation. Regarding the skills of the future, STEM is not the answer. The population does need to be computationally literate so the new technologies are not magic. But the most valued skills will be those that make us most human: skills like emotional and social intelligence, adaptability, and creativity.

3. Public perception.

The public's perception is driven more by Hollywood than reality. This has focused attention on very distant threats (like the fear that the machines are about to take over) distracting concern about very real and immediate problems (like the fact that we're already giving responsibility to *stupid* algorithms with potentially drastic consequences on society).

4. Industry.

The large technology companies look set to benefit most from the AI revolution. These tend to be winner-take-all markets, with immense network effects. We only need and want one search engine, one social network, one messaging app, one car-sharing service, etc. These companies can use their immense wealth and access to data to buy out or squash any startup looking

to innovate. Like any industry that has become rather too powerful, big tech will need to be regulated more strongly by government so it remains competitive and acting in the public good. The technology industry can no longer be left to regulate itself. It creates markets which are immensely distorted. It is not possible to compete against companies like Uber because they don't care if they lose money. Uber also often doesn't care if it breaks the law. As to fears that regulation will stifle innovation, we only need look at the telecommunications industry in the US to see that regulation can result in much greater innovation as it permits competition. Competition is rapidly disappearing out of the technology industry as power becomes concentrated in the hands of a few natural monopolies who pay little tax and act in their own, supra-national interests. For example, wouldn't it likely be a better, more open and competitive market place if we all owned our own social media and not Facebook?

5. Ethics.

There will be immense ethical consequences to handing over many of the decisions in our lives to machines, especially when these machines start to have the autonomy to act in our world (on the battlefield, on the roads, etc.). This promises to be a golden age for philosophy as we will be need to make very precise the ethical choices we make as a society, precisely enough that we can write computer code to execute these decisions. We do not know today how, for example, to build autonomous weapons that can behave ethically and follow international humanitarian law. The UK therefore should be supporting the 19 nations that have called for a pre-emptive ban on lethal autonomous weapons at the CCW in the UN. More generally, we will need to follow the lead being taken at EU on updating legislation to ensure we do not sacrifice rights like the right to avoid discrimination on the grounds of race, age, or sex to machines that cannot explain their decision making. Finally, just as we have strict controls in place to ensure money cannot be used to influence elections, we need strict controls in place to limit the already visible and corrosive effect of algorithms on political debate. Elections should be won by the best ideas and not the best algorithms.

6. Conclusions: The UK is one of the birthplaces of AI. Alan Turing helped invent the computer and dreamt of how, by now, we would be talking of machines that think. The UK therefore has the opportunity and responsibility to take a lead in ensuring that AI improves all our lives. There are a number of actions needed today. The UK Government needs to **reverse its position** in the ongoing discussions around fully autonomous weapons, and **support the introduction of regulation** to control the use and proliferation of such weapons. Like any technology, AI and Robotics are morally neutral. It can be used for good or for bad. However, the market and existing rules cannot alone decide how AI and Robotics are used. **Government has a vital responsibility to ensure the public good.** This will require greater regulation of the natural monopolies developing in the technology sector to

ensure competition, to ensure privacy and to ensure that all of society benefits from the technological changes underway.

7. Biography:

Toby Walsh is Scientia Professor of AI at the University of New South Wales. He is a graduate of the University of Cambridge, and received his Masters and PhD from the Dept. of AI at the University of Edinburgh. He has been elected a Fellow of the Australian Academy of Science, the Association for the Advancement of Artificial Intelligence and the European Association for Artificial Intelligence. He is currently Guest Professor at TU Berlin. His latest book, "Android Dreams: The Past, Present, and Future of Artificial Intelligence" is published in the UK on 7th September 2017.

5 September 2017

Andrew Ware, Dr Simon Beard, Dr Seán Ó hÉigeartaigh, Dr Shahar Avin, Martina Kunz, – Written evidence (AIC0150)

**Andrew Ware, Dr Simon Beard, Dr Seán Ó hÉigeartaigh,
Dr Shahar Avin, Martina Kunz, – Written evidence
(AIC0150)**

[Submission to be found under Dr Simon Beard](#)

Professor Kevin Warwick and Dr Huma Shah– Written evidence (AIC0066)

Professor Kevin Warwick and Dr Huma Shah– Written evidence (AIC0066)

[Submission to be found under Dr Huma Shah](#)

Warwick Business School University of Warwick – Written evidence (AIC0117)

Contributors:

Professor Mark Skilton, Professor of Practice (*pace of change & impact*)

Professor Juergen Branke, Professor of Operational Research & Systems
(*sectors to benefit*)

Dr Mareike Möhlmann, Assistant Professor (*sectors to benefit*)

Dr Emmanouil Gkeredakis, Assistant Professor of Information Systems
(*ethics*)

The pace of technological change

1. Work undertaken by Warwick Business School suggests the following changes:
 - The rate of sensors and AI algorithms automating speech, text and visual inputs is near >95% accuracy. In the next 5 years this will mean reliable human replacement of these tasks.
 - Creation of new AI related jobs in the fields of AI engineering and advanced data analytics will drive higher education research and the high-end skills job market.
 - 5% of daily tasks are completed through mobile services, self-service ordering and online payments. In the next 10 years, we predict this will move to 25% of today's job tasks. Tasks will be augmented by cyber intelligence devices and "assistants" in augmented reality (AR). We expect AR within 20 years will be widespread. Google recently cited as seeing 2030 as a potential timeframe for this to become a reality.
 - Superhuman tasks done by AI that no human could physically and/or cognitively do is already here, such as:
 - a. analysis of massive data sets for patterns
 - b. sub-one second manipulation of complex tasks that physically cannot be done by humans
 - c. advance space engineering and automated factories.
2. We predict in 10 years the emergence of *embedded intelligence* in the home, in transport and building devices. The need to plan for 20 years now will involve establishing social and technical platform projects that build these new capabilities. AI will transform the productivity of UK companies to allow them to compete with other countries.
3. We identified four key horizontal factors that will impede or accelerate productivity from using AI tools and systems, depending on how it is managed at policy and usage levels:
 - I. Cloud computing and network infrastructure access must be in place at the location to enable data and intelligence proxy to work. A counter argument is that smart processing on mobile devices will

- enable augmented intelligence in situ. However, this remains an unmet emerging market.
- II. Cyber security and protection of training data and algorithms to ensure they are used in a legal and commercial secure way.
 - III. Impact on energy and sustainability. Low carbon and efficiency improvements will be increasingly dependent on smart algorithms in connected appliances and buildings and vehicles to optimize energy usage and wastage.
 - IV. Ethical governance of algorithm usage.

Impact on society

4. The work conducted by WBS has identified key industries under threat from AI technologies taking-over human tasks.
 - Healthcare, cities, transport, digital delivered media, finance, insurance, legal education and agriculture sectors could be radically reshaped by AI replacing or augmenting many human jobs and tasks.
 - Some sectors such as media, publishing and banking could be completely disintermediated by AI automation processes. Governance and portfolio oversight might be the the only human job tasks remaining.
 - Creative industries could also be heavily augmented, but we expect polarization of both automation and non-automation in this sector, as it will remain largely driven by human ideas.
 - Supply chain, manufacturing and material extraction and movement would be driven by transformations in 3D printing.
 - Current limited 6DoF robotic degrees of freedom and general environment reasoning of situations will limit AI automation from entering directly into human social care, social interaction advisory services. This will remain the situation for several years until beyond 20 degrees of freedom robotics and higher degrees of manipulation enable general android like movement of robots in proximity of humans living spaces.

Industry

5. **Personalised manufacturing**

Industry 4.0 suggests a digital transformation of manufacturing resulting in smart factories and supply chains. Major changes to production and supply chains are anticipated as consumers demand more personalised products, moving away from the 'one-size fits all' approach to manufacturing. The ***biopharmaceutical industry*** will stand to benefit, with AI technologies being able to support the shift towards personalised medicine.

- Quality trumps cost in biopharma, embracing Industry 4.0 and AI provides a great offer to the sector, but quality will lie at the heart of the digital transformation.
- As an example, for autologous cell therapies if the starting cell concentration is lower than expected in the donor source, then AI technologies can be used, allowing the controls to make an informed decision on the increase in culture time required. Benefits will include reduced wasted, improved yields and possible more reliable patient-specific gene therapies.
- AI technologies will provide real-time visibility and control across complex cell and gene therapy supply chains from material sourcing through to manufacturing and temperature-controlled transport to patients.

6. Other manufacturing

Heuristics (or decision-rules) generated by AI technologies have been shown to enable factories to operate more efficiently than those run by man-made heuristics. AI driven heuristics have the potential to transform the manufacturing sector. However, we note some potential drawbacks/challenges:

- The rules generated by AI technologies are complex and humans are unable to understand how they work. This means humans are unable to learn from AI technologies that produce heuristics.
- Humans can model these and can see that they work in situ, but without understanding why they work they are unlikely to be 'accepted' by the 'mainstream'.
- AI driven heuristics is likely to threaten the more intelligent working-class groups in society.

7. Sharing economy companies

- With the rise of big data and networking capabilities, information systems can now automate management practices, performing complex tasks that were previously the responsibility of middle or upper management. These new practices, known as "**algorithmic management**," are the basis of ride-hailing platforms like Uber, with business models dependent on overseeing, managing and controlling myriads of workers who are not officially employed by the company.
- Algorithmic management practices are different from traditional management practices. They are characterized by the fact that workers behavior is constantly tracked, that workers performance is constantly evaluated, that they do interact with a 'system' but not with humans, and there is potentially low transparency about the underlying logic of algorithms.
- Work undertaken by WBS, identifies a series of mechanisms that Uber drivers use to regain their autonomy when faced with the power

asymmetry imposed by algorithmic management, including guessing, resisting, switching and gaming the Uber system.

- Sharing economy companies might need to consider:
 - showcasing detailed and illustrative feedback on the system to their workers.
 - finding ways for them to participate democratically in decision algorithms and policy to empower their workers.
 - develop a positive corporate culture, and even a feeling of identification among the workers.
 - approaches to support that preserve the human element. For example, a human customer support system.

Ethics

8. AI technologies work in strikingly opaque ways and act quite independently of their creators. The novelty of AI, especially novel machine learning techniques (e.g., deep learning), pertains to the capability of these technologies to act autonomously, without the guidance of humans. Unlike other technologies, the agency of AI is loosely entangled with the agency of its creators. As a result, fundamental ethical questions are raised about where moral agency and moral responsibility may lie. We perhaps need to consider that the **ethics of AI design** and the **ethics of AI use** are likely to be **decoupled**.
9. The **ethics of AI design** may be understood as follows: the extent to which AI engineers are more or less aware and educated about key ethical issues is likely to affect the (lack of) ethicality of their everyday design decisions. We may reasonably assume that resultant algorithms partly reflect the virtues, biases and (un)ethical or amoral intentions of their creators. For example, do AI engineers understand what data, which is used to train AI algorithms, is sensitive and in what ways? If not, what are the chances the resultant AI algorithms handle data with sensitivity to rights of privacy, etc.? Are AI engineers aware of their own implicit discriminations when designing AI technologies? Are their design decisions based on ethical recommendations made by institutes, such as the IEEE?
 - While the ethics of AI design have attracted a lot of attention recently, we believe more work is essential to investigate the influence of organizational contexts on AI design decisions. Most advanced AI technologies are developed in nascent organizations, often quite small start-ups. Many questions remain unaddressed: What kind of practices nurture ethical AI design? Are there more or less virtuous organizations that develop AI technologies? What does a virtuous organization that develops AI technology look like? Organization sciences may have a lot to say about what constitutes a virtuous organization. Yet, it is unclear how current research insights may be applicable in the AI context.
10. With the notion of **ethics of AI use**, we would like to highlight the

following, distinctive ethical dimension of AI: the best ethically educated AI engineers, with the best of intentions to avoid moral transgressions, may in fact have little control over the kinds of ethical decisions AI technologies will eventually make (Bostrom 2014). Even if AI engineers have a PhD in ethics, this cannot guarantee that the technologies they develop will behave and make decisions, which are ethical. An example might help illustrate this point.

- As we all know Google has developed sophisticated algorithms that choose which ads to display by taking into considerations what information its users input into its search engine. Harvard professor Latanya Sweeney uncovered the following: when you googled typical African American names, such as “Latanya Farrell,” you were shown ads offering to investigate possible arrest records. This result was not returned when searching for names, such as “Kristen Haring.”
- Google’s AI search algorithms, which were refined through feedback over time, were in effect making decisions with unintended ethical implications: people with certain kinds of names were associated with an ‘unethical’ past, without justification. As Luca, Kleinberg and Mullainathan (2016)¹³⁹⁶ reported: “people who searched for particular names were more likely to click on arrest records, which led these records to appear even more often, creating a self-reinforcing loop. This probably was not the intended outcome...”
- As the example indicates, even algorithms, which are deliberately designed to avoid discrimination, may inadvertently make ethically problematic decisions. The ethics of AI design may thus, in principle, be decoupled from the ethics of AI use.

11. AI technologies are largely exciting because they are endowed with autonomous agency, for example, to detect new patterns, modify original learned models and make novel decisions, which could not be anticipated by their designers. We need to recognize that, with such agency comes the capability to make morally-laden decisions, i.e., exercise moral agency. Beyond the major philosophical issues raised by this observation, we need to examine the pragmatic consequences. For example, can we hold designers, or their organizations, to account for inadvertent unethical decisions made by the AI technologies they created? Could boundaries be drawn between the moral responsibilities of designers and AI technologies? Or, are such boundaries, by definition blurred? Many more ethical questions remain to be asked.

12. We believe that ongoing empirical research is much needed to shed light on how, when, and why the ethics of AI design may be decoupled from the ethics of AI use. By doing further research, we would be able to determine whether it is possible to recouple ethics of AI design and use. For example, it may be that some decisions, which are at the moment left for AI

¹³⁹⁶ <https://hbr.org/2016/01/algorithms-need-managers-too>

technologies to make, could eventually be controllable through design. At the moment, however, it would be premature and unwise to make recommendations on how negative implications might be resolved.

6 September 2017

Weightmans LLP – Written evidence (AIC0080)

A response to the to the Select Committee on Artificial Intelligence Call for Evidence

Introduction

Weightmans LLP is an ABS and a top 45 national law firm with revenue of £95 million which employs 1,300 people across 9 offices. Weightmans is a full service law firm and is highly respected in the public sector, acting for many local, police and fire authorities, and NHS trusts. Weightmans provides strong, diverse commercial services for public sector bodies, large institutions, owner-managed businesses and PLCs and has a full family and private client service including: wills, tax, probate and residential conveyancing. Weightmans is a proud leading national player in insurance, with a formidable reputation and heritage with one of the largest national defendant litigation solicitor practices and an annual turnover in civil litigation work approaching £60 million.

Weightmans deals with motor, liability and other classes of claims for clients from the general insurance industry, other compensators including the NHSLA, local authorities, and self-insured commercial organisations such as national distribution and logistics companies.

Weightmans is actively involved in the insurance sector and has a number of major insurers as clients. Weightmans also specialises in the London Insurance Market, cyber liability, and automotive technology including autonomous systems and telematics, robotics and artificial intelligence, business crime, regulatory compliance, legal and commercial risk as well as offering in-house advisory services to insurers, non-insurer compensators and self-insureds.

In this response, we use the term 'artificial intelligence' (AI) in its broadest sense to refer to the development of technological systems which are able to perform tasks that would ordinarily require human intelligence. Within AI, there are several broad categories, such as machine learning, natural language processing, expert systems, speech/vision and robotics. These broad categories encompass techniques such as deep learning, predictive analytics, clustering and information extraction.

Our responses, unless otherwise indicated, apply to the area of professional legal services on which we are focusing our attention. More specifically, our current focus is on how we extract structured data from unstructured information and the application of legal reasoning and decision making as proofs of concepts. Once established in a particular legal domain, we envisage it will be possible to move these principles across domains. As a national firm, we anticipate that AI will impact on all aspects of our business, especially those that are transactional rather than advisory in nature. That said, even advisory work will be impacted, albeit in different ways. Our initial research into AI has seen us consider volume

litigation and property where the transactional nature of the work and the market suit automation and drive the need to reduce expense.

The call for evidence

We are pleased to be able to respond to your call for evidence as follows:

The pace of technological change

What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?

Technological developments in AI have grown massively over the last 40 to 50 years and legal AI has rapidly advanced. Historically, AI in the legal field has been focused on the academic pursuit of proving that a particular concept or system works (which in our view has looked, on occasion, like AI for AI's sake), usually using the same pool of tried and tested American jurisprudence as a form of control. In the legal sphere, there are a large number of start ups who are touting AI as being able to fill gaps in the market. Products are appearing in relation to due diligence and lease review and consideration is being given to process automation and machine learning at the lower value end of the claims process. Some aspects of the legal process are becoming more technologically assisted – 'court runners' would have traditionally lodged papers at The Rolls Building in London but from 27 April 2017, all claims must be filed electronically utilising CE-File.

There is an obvious synergy between law and AI and an overlap in relation to semi formal modelling - in law precedents, legal proof, rules and legislation. As Verheij has noted,¹³⁹⁷ modelling is a balance '*between the order of formal and the chaos of the informal. In law, rules have exceptions, reasons weighed, and principles are guiding. In AI, reasoning is uncertain, knowledge context-dependent, and behaviour is adaptive. This interest in the necessary balancing of order and chaos that is at the heart of both AI and Law points to the common subject matter that underlies the two fields: the coordination of human behaviour.*'

Over the next five years, we expect more law firms to:

- show an interest in AI;

¹³⁹⁷ Trevor Bench-Capon *et al*, 'A history of AI and Law in 50 papers: 25 years of the international conference on AI and Law' *Artificial Intelligence and Law* 20(3): 215–319, 2012.

- develop (if they have not already) innovation groups or boards; and
- in conjunction with commercial partners – some of the start ups referred to above - undertake proofs of concepts or utilise some of the ready made AI solutions that have been developed for the legal domain, by companies such as RAVN (recently acquired by iManage), Kira, Luminance or Rainbird. There will be others.

Over the next 10 years, the focus may be on applying the AI technology from one legal domain (such as contract review or due diligence) to other areas of law which may be less transactional in nature. Within the next 20 years, we see AI becoming more embedded in the legal profession and a change in the areas and types of legal services where human interaction and endeavours are required to be utilised – a successful application of legal AI will enable our lawyers to focus only on those elements of the work that they currently do that requires proper, formal, legal input. Law firms should be now including their response and approach to AI as part of their strategic considerations.

There is a concern in the legal industry that you may be better off as a technology firm that does law, rather than a law firm that does a bit of technology. The risk for traditional law firms is that technology firms that currently do no law enter the market and are well financed, data rich and data driven and without legacy issues and that they attempt to Hoover up subject matter expertise from traditional practices, who are not nimble enough to respond. We may see the rise of 'Uber Law' – technologically based, minimal overheads and able to act agilely in the market.

It does not appear that the pace of technological development will be the factor that hinders AI development, as there is evidence of legal AI being re-evaluated and applied to different legal domains. What may inhibit the application of legal AI is the cost of doing so, in terms of developing tailor made solutions or purchasing off the shelf products. Societal factors such as the impact on employment levels for example may also impinge on the speed of future AI development. Ultimately, any adoption of AI in the legal industry, be that by law firms or the courts, must ensure that the technology does not hinder access to justice for all members of society.

Is the current level of excitement which surrounds artificial intelligence warranted?

Whilst recognising that with the development of new technologies there will undoubtedly be an element of Gartner's hype cycle, we believe that the current level of excitement which surrounds AI is warranted.

We have spent the last year researching and embarking on proof of concept discussions with the ultimate aim of making one or more of the proofs of concept into a larger project. During this process, we have seen numerous

demonstrations and applications of AI that have been genuinely impressive, especially when compared with the human efforts that would be required to complete the same task. However, our excitement is tempered by what it may mean for employment and from a Government perspective, subsequent tax revenue.

At this point in time we are reminded in some way of the personal computer market in the 1980's. There were over a hundred different personal computers that were available to purchase. The market understood what all the pieces were but no one was sure how all of those pieces could best fit together. The market is also reminiscent of the dot.com boom in the 2000's with many start ups run by enthusiasts who believe AI will fundamentally change the world. Our view is that AI will be impactful but more prosaically than the evangelists would have us believe.

Impact on society

How can the general public best be prepared for more widespread use of artificial intelligence?

In this question, you may wish to address issues such as the impact on everyday life, jobs, education and retraining needs, which skills will be most in demand, and the potential need for more significant social policy changes. You may also wish to address issues such as the impact on democracy, cyber security, privacy, and data ownership.

We are of the view that there will be others better suited to respond to this question, save that cyber security, privacy and data ownership remain important legal principles (see our response to question 10).

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

We are of the view that there will be others better suited to respond to this question, however we note that the issue of data poverty may arise. This would result in a situation where wealthy companies who collect vast swathes of data would be able to turn that data into insights to generate further products or revenue. In a couple of years we may be in the situation where companies that don't collect and interpret their data struggle to compete. Following on from The Royal Society Report on 'Machine learning: the power and promise of computers that learn by example', more work on socio-economic insights appears to be warranted to mitigate the risks to the professions and to avoid training people (or placing them through apprenticeship schemes) for jobs that cease to exist in the next decades.

Public perception

Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?

Our initial response is yes, efforts will need to be made to raise public awareness but we are of the view that there will be others better suited to respond to this question.

Industry

What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?

In this question, you may also wish to address why some sectors stand to benefit over others, and what barriers there are for any sector looking to use artificial intelligence.

Whilst our response to this question is based on our own work in legal AI, we know that our insurance clients are investing in AI, robotics and automation to reduce the indemnity spend for their organisations. They are also utilising technology to improve the customer experience.

The potential applications for legal AI are vast and there are many benefits of AI utilisation for law firms (many of which apply to other areas where AI could be used) which include:

- increased efficiency,
- eradication of human error,
- redeployment of staff to more profitable tasks,
- risk mitigation – AI can review a much larger set of data than a human can and prevents any issues with random sampling,
- prevents having to redeploy staff from profitable tasks to less profitable tasks, for example to meet a deadline. For example, RAVN have reported that 800 hours of work was reduced to 40 by utilising RAVN Extract,
- improved profit margin and consequently could be more competitive in pricing,
- improved consistency.

However, AI may impact upon the continuing education of lawyers – if a machine can provide a quick answer, how will lawyers get practical education? This in turn may lead to a loss in skill set which may be required for alternative legal work. It will also potentially impact on the significant revenues the Government receives from law firms. As an example, Legal or Administrative Apprenticeships and the levy that the Government imposes on the scheme could come to an end as technology would take over these roles. Something needs to be done to future proof these roles and we may see an increase in different types of legal roles such as data engineers.

During the last year, we have been considering the legal reasoning side of legal AI to explore how AI can assist with decision making and we have also considered information extraction. As a key strategy partner to many different types of clients, we understand our responsibilities to take full advantage of technology to aid our service delivery solutions.

The recent online court hackathon, and the winning entry COLIN (the Courts OnLIN help agent) which utilised an Amazon Echo perhaps provide a glimpse as to the future of our Court system. Lord Justice Briggs' Final Report on the Civil Courts Structure Review published in July 2016 proposes an online court to deal with civil disputes of 'modest value and complexity' without the need to incur the cost of legal representation. The proposed online court will consist of three stages – issue/triage, alternative dispute resolution and determination. It has been proposed that the first stage will utilise a menu driven automated process and this seems to be crying out for the application of AI, in some form or another.

How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

We may be too late to address this. However, we are of the view that there will be others better suited to respond to this question, save that data security should remain an important overarching principle. In order to ensure a well-functioning economy, governance will be required in order to ensure that sufficient safeguards are in place to promote the behaviour and level of morality that society requires.

Ethics

What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?

In this question, you may wish to address issues such as privacy, consent, safety, diversity and the impact on democracy.

There are many overarching ethical issues surrounding the use and development of AI which apply to numerous sectors. These are complex issues that will require wider societal consideration and we list them merely to illustrate the range of ethical issues which includes, but are not limited to the following:

- how do we protect against unintended consequences?
- how do we ensure that the robots apply the right law appropriately, can they deal with applying empathy or accurately weighing up mitigating circumstances in the way that a judge can?
- how do humans “stay on top”?
- what happens if it all goes wrong i.e. where does legal liability lie?
- what will the impact of AI be on humanity?
- how do we prevent AI from applying bias on grounds such as gender, race or social class?
- will the greater application of AI result in unemployment? To some extent, this will in part be mitigated by a whole new raft of jobs that we had not even previously considered, assuming we have the education system that can provide the knowledge needed for the new roles.
- how accountable does AI have to be?
- do “robots” have “rights”?
- how do we guarantee security and prevent systems being hacked by those with malicious intent?
- how do we prevent the use of AI to circumvent security?
- how do we control the use of AI as a weapon?
- how do you keep a non tech route for those unable or unwilling (or no internet) to engage?
- how can inequality be dealt with – for example if companies become more profitable but with fewer employees?

In what situations is a relative lack of transparency in artificial intelligence systems (so-called ‘black boxing’) acceptable? When should it not be permissible?

Before AI is utilised, it should be thoroughly tested and indeed even when it is utilised, it will not often be at the exclusion of all human involvement. For

example, with technically assisted review of legal documents, this shifts the focus of the efforts of a lawyer from supervision of a junior fee earner or document review to the training and validation of an AI process. It is clear that lawyers will need to use AI in a non-negligent manner and this means adequate testing, validation and ongoing review when applications are in use. In some respects, this is no different to the requirement on a lawyer to keep themselves professionally up to date, or when supervising junior colleagues.

AI solutions that we have been investigating are auditable and would allow us to work backwards to understand if required the decisions that had been applied. This is perhaps more important in the legal domain as a law change could result in the re-engineering of an AI system so being able to show how a decision was reached and then to be able to see how the change could affect an ongoing claim is important.

The “black box” in some respects is what makes AI hugely interesting where a computer reaches useful conclusions or decisions that a human would not have been able to make (see move 41 of AlphaGo). Like all powerful technologies ultimately we have to build a regulatory framework that allows us to reap the benefits and limit the risks.

The role of the Government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

Funding is an issue that may inhibit the application of AI (notwithstanding that the technology is likely to continue to grow) and schemes such as Innovate UK should be adequately funded in future, and perhaps expanded to cover other areas. The Government should focus on looking at which roles will disappear with AI and fund transfers to new roles created by AI – we refer to our comments above in relation to apprenticeships. Traditional law firm models may struggle to get a sufficiently large enough budget to invest in AI when there is always something else that the budget can be put to, especially as it may be perceived by some as a speculative punt. We are aware that there is KTP funding to assist with innovate projects, but this process can be a bit cumbersome. More work should be carried out in joining up industries with those who will be utilising AI – in our sphere, this means lawyers engaging with technology companies and academia and to this end we were gold sponsors in June 2017 of the 16th International Conference on Artificial Intelligence and Law held in London.

Regulation of AI is important, but much will turn on how AI is defined – it may not be possible to have one type of regulation which encompasses all of AI. For example, Vermont in the USA has implemented a range of legislation to specifically deal with blockchain technologies. When determining the appropriate type of regulation, consideration will need to be given to:

- The adaptability and flexibility of existing legislation and legal framework - current data protection and data ownership principles may provide the legal framework, even if they do not always accurately match the technology in question.
- The proportionality of existing legislation and legal frameworks – will the regulation adequately promote innovation whilst at the same time protecting consumers / wider society.
- Who carries the risk of AI, in terms of product liability considerations. This may include consideration of whether AI, for example algorithms, be classified as a legal person or how agency principles extend to AI.

Learning from others

What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

We are of the view that there will be others better suited to respond to this question, but suggest that adequate research and consideration are given to the societal, legal and ethical issues and hurdles surrounding AI, so that they can be mitigated against in any future Government policy of programme of legislation.

Can we help?

If you have any questions or should we be able to assist you further in this regard, you can contact our Innovation Group:

Stuart Whittle

Partner

Rob Williams

Partner

Dr Catriona Wolfenden

Solicitor

Weightmans LLP

5 September 2017

Wellcome Trust and the Association of Medical Research Charities (AMRC) – Written evidence (AIC0202)

7 September 2017

Key points

- Artificial Intelligence (AI) technologies have the potential to deliver improvements in clinical decision-making, patients' healthcare and biomedical research.
- Clear oversight is needed for AI developments to be used for the benefit of society. This should cover both the inputs into AI systems, particularly with regard to people's personal data, and the outputs or decisions they produce to ensure they are fair and unbiased, with accountability for decisions made.

Introduction

1. We are pleased to respond to this inquiry on artificial intelligence. We strongly support Parliament's interest in this area and it is a positive step that conversations about the potential and value of AI in society are being promoted by the Committee. Given Wellcome's and the AMRC's positions as research funders, our response focuses on the potential and impacts of AI on healthcare and biomedical research, and its potential social, cultural and ethical implications.

Definitions

2. We take 'Artificial Intelligence' to be an umbrella term covering a number of functions, including machine learning, deep learning, image recognition, natural language processing, computer vision and robotics. Our response focuses primarily on machine learning.
3. The term 'Artificial intelligence' is not particularly helpful: it is usually used to denote systems that are capable of self-learning, but to describe these as 'intelligent' may be misleading. The term also generates sensationalist perceptions of what these technologies are capable of, which do not help public discourse about their potential and how they should be used.
4. The current excitement about AI largely results from changes in the way machine learning algorithms can be combined together and trained. These provide tools for extracting meaning from large quantities of data in a way that was not previously possible.

The pace of technological change

5. The pace of technological change is currently extremely fast. This is being driven mainly by the recognition of the potential commercial benefits that AI can create. As is common with expectations about new technologies, the excitement may be overblown at this stage. Nonetheless, there are key areas in which there are tangible developments with real potential:
 - In the health sector, pharmaceutical companies are currently interested in AI to replace or inform drug discovery pipelines¹³⁹⁸. Wide-ranging disease databases can also be mined by researchers for previously undetected patterns, for example to identify potential new cancer drug targets¹³⁹⁹. AI is also looking promising for enhanced disease detection via better image recognition, and for analysing the vast quantity of new data being produced by genome sequencing efforts, which may help further our understanding of disease and assist with earlier diagnosis.
 - In terms of basic research, machine learning is helping to further our understanding of how the brain processes information from the outside world, for example, modelling of deep neural networks in visual object recognition.
 - For healthcare delivery, the realistic implementation of AI in the next few years will likely be in the areas of app development for diagnostic purposes and the implementation of personalised support and treatment regimes. For example, Arthritis Research UK has developed an AI-based virtual assistant in conjunction with IBM to for people with musculoskeletal conditions to access information to support self-management.¹⁴⁰⁰ For those with Parkinson's Disease, there have been promising advances using Extended Reality (ER). This involves a headset overlaying virtual reality cues onto the user's surroundings to aid with walking.¹⁴⁰¹
 - AI-based clinical decision support tools are also likely to complement (not replace) clinicians' expertise. These could lead to medical decisions being informed by better data, enable data collection to be better

¹³⁹⁸ For example, <http://benevolent.ai/news/articles/this-ai-unicorn-is-disrupting-the-pharma-industry-in-a-big-way/>

¹³⁹⁹ Institute of Cancer Research, CanSAR database [accessed 25 August 2017] <https://cansar.icr.ac.uk/>

¹⁴⁰⁰ Arthritis Research UK: Virtual Assistant [accessed 25 August 2017] www-03.ibm.com/press/uk/en/pressrelease/51828.wss

¹⁴⁰¹ Uses for Extended Reality in Parkinson's Disease, Northeastern University, Boston www.northeastern.edu/rise/presentations/augmented-reality-parkinsons-assistive-tool-based-visual-cues/

harmonised and thus form a virtuous circle in improving the data for further machine learning development.

6. Access to good-quality, comprehensive, standardised datasets is critical for the development of AI-based tools and technologies. As indicated in the recent Life Sciences Industrial Strategy¹⁴⁰², the NHS is a fantastic potential resource but is not yet equipped to capitalize on the data it collects. There are also significant problems of public confidence and trustworthiness in the system for using patient data, particularly where commercial interests may be involved. These factors will inevitably hinder the development of AI in healthcare particularly.

Impact on society

7. It is currently difficult to separate out hype and speculation from realistic predictions about the potential impact on society of AI-based technologies. The issues the Committee raises in relation to societal impacts all merit substantive research and discussion but it is worth noting that many of them are not necessarily driven by developments in AI: the technologies themselves are tools and we face societal choices about how and to what ends they should be used.
8. There is a risk of perpetuating health inequalities or creating lack of equity in access to healthcare systems if the impact of AI innovations is limited to those with digital skills and access, for example to smartphone apps.
9. There are social, cultural, legal and ethical questions about the role, value and potential of AI in biomedical research that have yet to be researched. To develop a clear understanding of these issues and how to address them, independent academic research along with public engagement and involvement at an early stage in the development of AI tools is crucial.

Public Perception

10. If AI technologies are to be used to benefit society, it is extremely important to engage with the public about: what these technologies can potentially offer; what they require (in terms of data); how they could be used; what the risks are; and what rights people have in relation to their data that could be processed via automated means¹⁴⁰³.

¹⁴⁰² Life Sciences Industrial Strategy [accessed 1 September 2017]

www.gov.uk/government/publications/life-sciences-industrial-strategy

¹⁴⁰³ This will be essential in light of the EU General Data Protection Regulation right to restrictions on automated decision-making (Article 4(4)).

11. A core part of the Understanding Patient Data (UPD) initiative concerns horizon scanning about new digital technologies such as AI and considering how best to support public conversations about the use of such technologies in healthcare and research¹⁴⁰⁴. In Autumn 2017 UPD is partnering with the Academy of Medical Sciences to undertake a public dialogue with patients, clinicians and publics to explore attitudes towards the use of new digital technologies in healthcare and research. We would be pleased to update the Committee on this work when complete in early 2018.

Ethics

12. Some ethical implications of AI are well-established because they relate to the ethics of the use of 'big data' more broadly¹⁴⁰⁵, as large datasets are required as inputs to train algorithms. The increasing capacity to interrogate large, complex and better-linked datasets has implications for privacy and the risks of re-identifying individuals from anonymised datasets. This privacy issue is not unique to AI but the capacity of AI to process vast quantities of data may magnify it. There are also questions over the extent to which consent to allow data processing by a machine learning algorithm can be truly informed.
13. However, there are wider ethical implications of AI that go beyond privacy and consent, such as ensuring people do not get discriminated against or harmed as a result of decisions made via AI. Managing these will require both good governance and oversight, and addressing wider questions about equity of access to healthcare¹⁴⁰⁶.
14. These ethical challenges arise in part because of the lack of transparency afforded by some forms of AI: so-called 'black-box' processing whereby it is not possible to explain or account for how an output is reached. This means that potential biases or errors in outputs may go undetected and may lead to discriminatory decisions and unintended consequences. This also raises substantial questions of accountability as it is unclear who is responsible for the decisions or outputs of an algorithm.
15. Careful consideration should be given to the balance between the accuracy of an algorithm and the degree of transparency over its process, allowing

¹⁴⁰⁴ Understanding Patient Data [accessed 1 September 2017]
http://understandingpatientdata.org.uk/sites/default/files/2017-08/New%20Tech%20Slide%20Deck_2.pdf

¹⁴⁰⁵ Wellcome Trust response to House of Commons S&T Committee on 'The Big Data Dilemma', p.5: <https://wellcome.ac.uk/sites/default/files/wtp059804.pdf>

¹⁴⁰⁶ Presentation by Prof. Mike Parker (University of Oxford):
<http://understandingpatientdata.org.uk/sites/default/files/2017-08/Parker%20Patient%20Data.pdf>

the impact of new technology to be captured while maintaining rigorous safeguards. A diverse workforce of developers and a wide range of stakeholders involved in the creation and regulation of AI technologies may help also mitigate the risks of introducing bias into AI systems.

16. Decisions about the risks inherent in a lack of transparency will need to be made on a case-by-case basis, within a framework of governance and oversight that is guided by clear ethical principles.

The role of the Government

17. Government has a role to play in two key aspects of the development and use of AI: protecting people's rights over data about them (as inputs into AI systems), and ensuring new technologies are used for the benefit of society (as AI outputs).
18. AI technologies critically depend on the data available to them for training and development: in the health sector, the largest and most useful datasets are likely to be population-level data gathered through the course of clinical care. These datasets have tremendous potential value for AI, but they comprise people's most sensitive information and must be carefully protected. The new data protection law includes provisions on automated decision making, which could be delivered through AI, and will play an important broader role in updating privacy rules. However, it is important that law is kept under review to keep pace with changes in technology.
19. In terms of the outputs of AI, proportionate and appropriate new approaches to regulating self-learning software are yet to be developed. This needs to consider transparency on the division of decision-making responsibility between humans and AI systems and to ensure there is appropriate accountability for decisions made. The focus of any regulatory framework for AI-based products should be sector-specific and designed for the context and purposes for which the technologies would be used, for example as diagnostic tools, rather than on the technologies themselves. Any regulatory efforts need to be informed by technical and ethical expertise as far as possible and kept up to date as the field evolves.
20. Discussions about the recommendation of a stewardship model of data governance from the Royal Society/British Academy¹⁴⁰⁷ and the Nuffield Foundation's proposals for a Convention on Data Ethics¹⁴⁰⁸ are

¹⁴⁰⁷ Royal Society & British Academy: Data Governance [accessed 15 August 2017]

<https://royalsociety.org/topics-policy/projects/data-governance/>

¹⁴⁰⁸ Nuffield Foundation press release [accessed 3 July 2017]

<http://www.nuffieldfoundation.org/news/nuffield-foundation-announces-additional-%C2%A320-million-research-funding-fellowship-programme-and->

progressing. In this context it will be extremely important for Government to set out a clear vision for what role and function it will seek to develop for the oversight, governance and regulation of new digital technologies, including AI. Any new body/bodies will need to be flexible and responsive to new developments; be able to identify regulatory gaps and challenges as they arise; and convene expertise to build consensus on what appropriate oversight and governance looks like. The role of existing regulators, such as the Medicines and Healthcare products Regulatory Agency and the Information Commissioner's Office, should not be duplicated by new bodies, and these agencies must be joined up with this forward thinking approach.

Learning from others

21. AI is a rapidly developing field and it is inevitably difficult for policy and regulation to keep pace. We are not aware of other countries or regions that have already developed a systematic approach to the issues raised in this inquiry, although there is significant interest and activity in this space globally^{1409, 1410}.
22. In terms of policy approaches, the challenge will be to ensure fundamental discussions about the ethics, policy and governance of these technologies are joined-up across diverse sectors and happen early enough to influence the direction of their development. Useful insights from sectors that may be further ahead (for example, autonomous vehicles) can inform the development of approaches in other sectors. This includes lessons learned where difficult ethical and social questions might arise.
23. The UK has an opportunity to take a global leadership role on policy, regulation and establishing the right ethical frameworks for the use of AI to benefit society. This would build on the country's good reputation for managing and regulating complex and ethically challenging technologies, for example, on mitochondrial donation techniques.

7 September 2017

¹⁴⁰⁹ Centre for the Fourth Industrial Revolution [accessed 15 August 2017]
<https://www.weforum.org/center-for-the-fourth-industrial-revolution/areas-of-focus>

¹⁴¹⁰ Partnership on AI [accessed 15 August 2017] <https://www.partnershiponai.org/>

Vishal Wilde – Written Evidence (AIC0004)

Author Bio: Vishal Wilde FRSA is writing in a personal capacity and his views do not necessarily reflect any of the organisations he is affiliated with. He is on the list of approved parliamentary candidates for the Liberal Democrats and is an incoming Civil Service Fast Streamer (on the Generalist scheme). He writes on economic, political and financial topics as a Featured Columnist for The Market Mogul (where he is also an Editorial Associate). He has written for think tanks such as The Cobden Centre, the Center for a Stateless Society (C4SS) and the Adam Smith Institute on a broad variety of topics. He is also Co-founder and Chairman of Project Shanthistan, a very nascent think tank and movement which seeks to foster peace, prosperity and cooperation in South Asia with an eventual aim of unification through promoting peoples' social, political and economic freedoms.

At the time of submission, he is in the final stages of studying for an MSc in Advanced Computer Science with Internet Economics at the University of Liverpool and holds a BSc (Hons) in Philosophy, Politics and Economics (Economics major) from the University of Warwick.

Impact on Society

Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?

- The large tech firms (and those who can afford to invest in developing, deploying and purchasing tech for these purposes such as financial services firms, certain law firms, some professional services firms etc.) with access to the subsequently mentioned 'data-based monopolies' and who are in a position to take advantage of the legal, deterrent-based and excessively punitive nature of intellectual property regimes.
- As such, the productivity gains associated with the rapid development and increasingly sophisticated and ubiquitous deployment of various forms of AI are more concentrated within particularly large, well-endowed firms (in the tech industry, this includes firms such as Facebook, Google, Microsoft, Apple, IBM etc.) but which does proliferate in some (albeit limited) capacity to SMEs and micro-enterprises also. This is particularly because investing in AI can be (and often is) financially costly, time-intensive and risky. In this sense, it could be a potential contributing factor to the 'Productivity Puzzle' that is afflicting the developed world and especially Britain within a changing Europe. After all, Copyright and IP affected a far smaller and less integral part of the economy during the recovery periods

of pre-2008 recessions – now it affects vast swathes of vital industries.

- Those who live in rural areas and who do not have access to broadband also do not feel the benefits of the burgeoning internet economy and the productivity gains associated with AI nearly as much as other parts of the country with satisfactory internet access and digital capabilities.
- Disparities can be mitigated through reforming intellectual property regimes broadly (especially Copyright as it relates to software) and more rationale for this will be provided throughout the submission. We need to stop affording corporations the legal protection of Copyright regarding their code in particular since this acts as a simultaneous legal deterrent and barrier to entry/establishment in the industry. James Bessen (Boston University School of Law) and Eric Maskin (Nobel-Laureate and Professor at Harvard) published a paper entitled "Sequential innovation, patents, and imitation" in 2009 in the RAND Journal of Economics which argued that "society and even inventors themselves may be better off without such protection). Indeed, although these researchers refer to patents, it still falls under the broader scope of intellectual property regimes. For the benefits of AI to be felt across all industries, patent protection, Copyright and intellectual property regimes more broadly across all industries need to be significantly liberalised and/or thoroughly reformed.
- Relating back to the Productivity Puzzle, the difference between this recession and the recovery periods after previous recessions is that Copyright affects far more substantial parts of the economy than it previously ever has (due to the growth of the internet, proliferation and ubiquity of software, usage of it etc.). This is another potential interpretation/narrative for the Productivity Puzzle. Indeed, the influence of technical progress upon productivity is well-established (both historically and contemporarily).
 - Indeed, due to the associated, exceedingly high supernormal profits associated with the monopoly privileges granted by legal IP, this can conceivably to a corresponding exacerbation of input costs (in anticipation of higher profits) and, thus, it artificially constrains growth of the industry and its associated innovative capabilities.
- Substantial human capital is necessary – the deficit of STEM graduates in the UK is well-known and, thus, this is an opportunity to reconsider government policy on Higher Education. I would particularly recommend liberalising the Student Loans Company's policies so that those who borrow tuition-fee and maintenance loans are able to spend them abroad and are not merely confined to spending them in UK Higher Education institutions.
 - To begin with, UK students should have the ability to study in other countries (where education can often be less expensive than in the

UK) and, thus, it would improve accessibility to higher education.

The idea that they will simply choose to reside abroad is a fallacious argument because the alternative is that they would simply leave to better (economic) conditions after having been educated here.

- Secondly, those students who do study at foreign universities that charge less than domestic institutions would graduate with far less student debt which would help reduce the burden upon both taxpayers and students. Students would also return with a diversity of skills alongside knowledge of various cultures and languages (linguistic capabilities amongst British students being a comparative weakness in an increasingly 'globalised' world).
 - Since many UK students could study abroad, UK universities could afford to admit more international students and re-invest the surplus from fees into research capabilities (especially since universities' research funding is under significant threat due to Brexit: knowledge- and innovation-centres are vital for Artificial Intelligence and the economy's productivity and growth more broadly).
 - Expanding access to STEM subjects (amongst others also) in this way will be especially useful but we must also ensure that the gendered, ethnic and racial gaps associated with the uptake of STEM subjects are not neglected in the process (for, if they are, this would exacerbate pre-existing inequalities in societies and reinforce animosity and social divisiveness). Most importantly, those from low-income households and disadvantaged socioeconomic backgrounds must be assimilated into a general increase in the accessibility of STEM subjects as well as Higher Education more broadly to enable social mobility. Indeed, allowing people to spend their student loans abroad in countries where tuition fees can often be less expensive (such as continental Europe, Asia, Africa and South America) will work to encourage people from low-income backgrounds to engage in higher education.
 - Even students from high-income backgrounds could use these loans to subsidise their study in North America and Australia (for example) and the places they vacate at British universities could be filled by international students here (which would, in turn, improve UK universities' financial capabilities).
 - Indeed, one needs only a rudimentary acquaintance with the culture and diversity of Silicon Valley and the wide variety of educational and professional backgrounds in the tech industry there to discern how this could thoroughly benefit Britain in these extremely uncertain times.
- There is also a need to ensure talent is available to firms: this means ensuring consistently liberal immigration laws and straightforward immigration procedures. If Brexit occurs in the 'hardest' form possible, this

would require enabling far freer movement of (skilled) labour with the rest of the world whilst simultaneously ensuring streamlined immigration procedures for skilled labourers from the EU also.

- At the very least, if even the physical presence of migrants is politically infeasible due to elected representatives' and the government's collective inability to act in the best interests of peoples, there should be arrangements in place to make working with, employing and (sub-)contracting foreign-domiciled workers as simple and as non-onerous as possible.
- A liberalisation of land-use restrictions is necessary to alleviate intra-regional disparities across the UK. Indeed, it is no coincidence that a paucity of diversity of industries in rural and semi-rural economies means that there is a relative diminution of income and opportunities available to rural and semi-rurally domiciled populations. The increase in investment into infrastructure and industries in these previously restricted domains (which would be enabled through the liberalisation of land-use restrictions) as well as the accompanying increase in incomes will naturally incentivise greater investment into internet accessibility and capability in these areas (which would help alleviate those intra-regional inequalities within the UK).
 - Indeed, there should also be an accompanying liberalisation of what properties can be used for 'commercial'/'research' purposes to take full advantage of increasingly flexible labour market institutions (which, again, could help feedback into alleviating the UK's Productivity Puzzle from a Human Resources perspective).

Industry

How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?

- This also relates back to the suggestions made within the 'Ethics' section but significant reasons for the 'data-based monopolies' of large corporations are Copyright and Intellectual Property regimes.
- It is time-intensive, costly and risky to develop AI systems that can harvest, learn from and make profitable/welfare-enhancing use of data. One of the massive difficulties is hiring talent since the asymmetric information in the market introduces a significant potential for and degree of adverse selection and moral hazard when trying to find suitable developers. As such, it is even more risky for smaller firms to hire talent that was previously at a large corporation.

- If programmers/AI developers can show their capabilities without being restricted by the Copyright afforded to corporations over their work, this will significantly increase their bargaining power in the industry (and anyone with even a basic acquaintance with labour markets will be aware that bargaining power determines workers' wages). Thus, as programmers' wages increase, this will also encourage an increase in the supply of AI developers and more people will take up STEM subjects when they see this corresponding increase in wages (thereby alleviating the talent shortage). As such, Copyright and Intellectual Property regimes more broadly work to suppress wages in the industry and also inhibits the mobility of programmers within and across industries.
- Furthermore, the patents in other industries (such as in pharmaceuticals and biotech, which are increasingly impacted by developments and advances in AI) unduly inflate the costs associated with innovation throughout the economy. This subsequently constrains productivity growth through inhibiting the diffusion of technologies throughout society.

Ethics

In what situation is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?

- Companies should, as a matter of principle, disclose the variables they collect data on for each individual and to each individual for whom it is collected. They need not disclose the actual data itself (since part of the terms of use of the services of these companies is that data can be collected and used) but there is the significant matter and concern of *informed consent*. That is, one needs to be aware of *what* they are consenting to (what variables on which data are being collected). Nevertheless, companies continue to profit from asymmetric information in markets (and, often, by harvesting and using data that has only been implicitly consented to and may not necessarily be explicitly consented to). Furthermore, there is the practical worry of being unable to foresee the unintended consequences of AI and, if people do not know what parts of their data are being used by AI systems, they will be unable to know what aspects of their personality are being catered to, responded to or even manipulated by AI systems (unintentionally or otherwise).
 - It may be tempting for some to have governments mandate this but a more effective and holistic 'fix' (in both the medium- and long-term, such that industrial growth and productivity is not unduly constrained) to this problem may be associated with the aforementioned liberalisation (preferably abolition) of Copyright and significant liberalisation and/or reform of IP more broadly. This

- could have several incentive effects that correct these highly unusual and legally-privileged information asymmetries derived from data-based monopolies.
- Intellectual Property (Copyright in particular in the case of software) can actually significantly inflate the costs associated with setting up new social networks and social media. Thus, appropriately reforming and liberalising IP will work to significantly reduce the barriers associated with entry into and establishment within these industries.
 - With a potential (further) proliferation of social networks and social media with comparable capabilities to incumbents, each company would be able to offer different degrees of privacy, transparency etc. to heterogeneous consumers/users with diverse preferences. Thus, the desire for different standards of privacy, morality and transparency would more likely reflect the diversity of moral values in society regarding data.
- The most obvious argument for transparency is where national security is at risk. Indeed, with an ever-growing portion of economic infrastructure being significantly (often even wholly) reliant upon cyber infrastructure, this is a major vulnerability in the economy which can be exploited by malicious actors (these actors being of both the state and non-state variety). There may be arguments made for national security institutions (such as the military) being afforded access and demanding transparency where they deem it to be necessary for national security. However, the objectives of these demands can also be met through significantly liberalising and/or reforming IP regimes instead.
 - To begin with, intellectual property can stifle information security innovation.
 - If corporations are accustomed to the idea that the government will legally defend their inventions and innovations (and, thereby, deter innovation by smaller, disadvantaged, would-be competitors) this inherently privileges incumbents over entrants and it also discourages innovation in the domain of information security (pertaining to human and cyber factors or even otherwise).
 - Reforming IP and Copyright in particular means that there will be an authentic institutional incentive to significantly and constantly improve information security which would greatly improve the resilience of an increasingly digitally-reliant economy.
 - Even where there are companies that solely rely on IP for their business models, they should be responsible for their own protection of IP (if they deem it necessary) rather than relying upon legal-deterrents, legal-enforcement and government protection.

The role of the Government

What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?

- No, artificial intelligence should not be regulated – even if the government wanted to, it is exceedingly difficult to see how it should be and how it could be (especially since technical expertise in the field is scarce enough as it is). Government should refrain from dampening one of the few enablers of productivity growth in the developed world. Indeed, it is a well-established historical and contemporary fact that technical progress is a key engine for productivity growth.
- Liberalise and/or significantly reforming intellectual property regimes (ideally scrapping Copyright pertaining to software entirely) to reduce the cost of knowledge-diffusion and subsequent innovation throughout society.
- Ensuring a consistently liberal immigration regime – the importance of migrant labour for tech cannot be understated: one needs only a cursory acquaintance with Silicon Valley in California to understand this.
- Liberalise the Higher Education market through allowing students to use their borrowed funds for educational purposes from the Student Loans Company abroad as well as domestically to invoke the aforementioned benefits (such as improving accessibility for poorer students, reducing student debt, reducing taxpayer burden and potentially improving universities' financial circumstances in these extremely uncertain times).
- Liberalise land-use restrictions to not only alleviate the housing crisis and to also formally allow commercial activity to take place in a variety of properties (to take full advantage of the potential for a truly flexible labour market in increasingly flexible contexts) rather than institutionally incentivising its occurrence in certain types of properties. Indeed, a liberalisation of land-use restrictions would correspond to investment in infrastructure in rural and semi-rural areas which would help diversify the economies there, help alleviate intra-regional disparities within the UK and also naturally incentivise financial investment into the infrastructure necessary to ensure adequate internet access to and fast broadband capability for 100% of the UK population.

Learning from others

What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?

- None (or, to be more generous, very few). This is a relatively new frontier and the UK should take the lead in making a revolutionarily innovative climate for the field and the associated industries that are impacted by/benefit from it. Indeed, excellent artificial intelligence research is not confined to particular regulatory regimes or regions but, rather, they are dispersed across the world according to the quality and concentration of (intellectual) human capital as well as investment capabilities and infrastructure.
- To look at other countries as models of how the UK *should* be in this instance may inadvertently constrain future innovation and corresponding (productivity) growth.

26 July 2017

Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos and Professor Austin Tate – Written evidence (AIC0029)

Professor Chris Williams, Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos and Professor Austin Tate – Written evidence (AIC0029)

[Submission to be found under Professor Robert Fisher](#)

Professor Rebecca Williams – Written evidence (AIC0206)

I write to provide evidence from my research which may be of interest in addressing points 3 and 10 of your call for evidence, concerning the questions of how society may best be prepared to deal with the challenges arising from Artificial Intelligence (AI) and Machine Learning (ML) and the role of the government in that. My research focuses on how existing rules of criminal law will need to adapt to deal with ML and how liability for damage caused by ML might best be structured in the criminal context.

The law has always struggled with finding a successful method to attach criminal liability to situations involving more than one human being. How are concepts such as mens rea to be attached to groups of individuals, rather than single individuals? And proving that the relevant events and thought processes took place in a commercial context has always been difficult, even given the possibility of dawn raids on business premises. But the advent of ML in business seems likely to pose further challenges on both these fronts, of a kind which suggests that we will need substantially to change the way in which we think of detection and culpability in the 21st Century. If a ML system decides that it can best achieve its objectives by defrauding regulators, or by endangering or injuring persons or property,¹⁴¹¹ how should the law respond?

Establishing culpability

As far as establishing culpability and mens rea are concerned, the law has adopted various approaches.¹⁴¹² Some of these, such as the personal liability of corporate directors,¹⁴¹³ or the 'identification doctrine', which seeks to identify the 'controlling officers' who are the 'directing mind and will' of the company,¹⁴¹⁴ vicarious liability¹⁴¹⁵ and statutory liability for specific corporate officers¹⁴¹⁶ attempt either to identify certain individuals who can be made personally culpable for the relevant wrongdoing, or who can be identified with the company for the purposes of making the latter liable, in an attempt to keep corporate criminal liability within the usual orthodox framework of criminal law as it applies to human beings. But if the 'directing mind and will' in a particular case is no

¹⁴¹¹ For examples of the existing capacity of even quite basic AI or ML systems, see, e.g. http://lesswrong.com/lw/mrp/a_toy_model_of_the_control_problem/ or <http://www.popsci.com/scitech/article/2009-08/evolving-robots-learn-lie-hide-resources-each-other> . For further discussion see N Bostrom, *Superintelligence*.

¹⁴¹² See further Smith and Hogan's *Criminal Law* (14th ed 2015) Ch 10.

¹⁴¹³ Including via s 8 of the Accessories and Abettors act 1861.

¹⁴¹⁴ See further *Lennard's Carrying Co Ltd v Asiatic Petroleum Co Ltd* [1915] AC 705, 713 per Viscount Haldane LC. See also *JF Alford Transport Ltd* [1997] 2 Cr App R 326 at 331 and *Tesco Supermarkets v Natrass* [1972] AC 153.

¹⁴¹⁵ *Moussell Bros v London and North-Western Rly Co* [1917] 2 KB 836.

¹⁴¹⁶ E.g. s 12 of the Fraud Act 2006.

longer that of a human being, even these attempts seem likely to fail. Indeed, not only will it prove even more difficult to find mens rea on the part of a human being in the company, it will not even be possible to prove causation, on the basis that the directors and even the programmers may be able to argue that there has been something akin to the 'free voluntary action' of a third party which is recognised by the criminal law as breaking the chain of causation and preventing the ascription of liability to the defendant.¹⁴¹⁷ And this of course provides a great incentive for human agents to avoid finding out what precisely the ML system is doing, since the less the human agents know, the more they will be able to deny liability for both these reasons.

Detecting Culpability

Not only will it be difficult to attach liability for any wrongdoing to a human being, even the detection of that wrongdoing in the first place will become more difficult with the introduction of autonomous decision-making systems. Even the more consequentialist analyses of criminal law which focus on achieving deterrence of wrongdoing rather than attributing blame for existing culpability often focus on incentivising 'gatekeepers' in the corporate system, again focusing on the control of specific human beings with access to particular knowledge about the workings of the firm, such as lawyers, underwriters, outside directors or accountants.¹⁴¹⁸ However, the development of ML systems for use in business may mean that the 'controlling 'mind and will'' in a particular case is no longer that of a human being. It is clear that both Volkswagen's 'defeat device' used to cheat admissions tests¹⁴¹⁹ and Uber's use of 'greyball' technology¹⁴²⁰ to identify law enforcement agents were the result of conscious decisions by human beings, but this will not necessarily always be the case. It is by no means impossible that a ML system instructed to maximise the profitability of a particular operation could figure out that profitability decreases with the imposition of fines and thus decide to maximise profitability and decrease fines by finding its own way to evade law enforcement.¹⁴²¹ And were it to do that, the process would be very difficult to detect. There would be no individual with a guilty conscience to experience the prisoner's dilemma. Even in cases such as Volkswagen the number of human beings who need be involved in the wrongdoing is much smaller than might previously have been necessary, but if the decision is taken by ML that number may decrease to zero. So the first indication of a flaw or 'wrongdoing' in the decision-making process would be the existence of external harm, such as an increase in vehicle emissions or the pollution of a river. And even if the regulator or law enforcement agency did suspect or monitor some

¹⁴¹⁷ *R v Kennedy (no 2)* [2007] UKHL 38.

¹⁴¹⁸ R Kraakman, 'Corporate Liability Strategies and the Costs of Legal Controls' (1983-84) 93 *Yale Law Journal* 857, see also J Arlen and R Kraakman, 'Controlling Corporate Misconduct' [1997] 72 *New York University Law Review* 687, fn 24.

¹⁴¹⁹ See further <https://www.epa.gov/vw>.

¹⁴²⁰ https://www.nytimes.com/2017/03/03/technology/uber-greyball-program-evade-authorities.html?_r=1

¹⁴²¹ See above, n 1411.

harm of this kind, it is difficult to imagine what would be the equivalent of a 'dawn raid' on an algorithm. Even if manufacturers were prepared to grant 'backdoor' access to their software (which would be a security risk about which they could legitimately have concerns), would it even be possible to 'watch' the algorithm cheating the system? It is relatively simple to observe the choices made by the toy control experiment robot,¹⁴²² but once decisions are being taken in the real world, the reasons for particular choices may be much less obvious, not least because programmers might quite legitimately argue that they have used code obfuscation to protect themselves from competitors.¹⁴²³ One option might be to require systems to log their decision-making processes carefully, but even that might simply lead to such large quantities of complex thought processes that it would be unrealistic to expect any kind of law enforcement agent to sift through them.¹⁴²⁴

Reframing liability

How, then are we to detect wrongdoing and incentivise compliance and harm reduction in the light of these challenges? How do we need to modify our traditional framework to ensure that the criminal law can still do its job properly?

The most obvious starting point might be to prohibit the use of ML in certain contexts,¹⁴²⁵ but as Sandberg notes, 'there will be a tension between law-abiding and capable design'.¹⁴²⁶ There may be circumstances in which we forgo the potential benefits of ML autonomy in order to reduce their risks but (a) there will be circumstances where we 'accept unverifiable but useful autonomy'¹⁴²⁷ and (b) it is not clear that we would be making the most beneficial choice if we were to do so.¹⁴²⁸

One standard option in circumstances where proof of mens rea is difficult is to abandon the requirement altogether and opt for an approach of strict liability¹⁴²⁹, in which the prosecution need prove only the relevant harm, not that the defendant had any particular mental attitude in relation to it. But this is problematic in one of two ways. Either the resulting offence is 'not truly criminal'

¹⁴²² Above n 1411.

¹⁴²³ See, e.g. <http://blogs.adobe.com/acrolaw/2011/06/code-obfuscation-for-patent-and-court-filings/>.

¹⁴²⁴ For discussion in the civil, rather than criminal context, see Anders Sandberg: <http://www.oxfordmartin.ox.ac.uk/opinion/view/340>.

¹⁴²⁵ Yoshua Bengio of the University of Montreal has argued for an outright ban on the military use of AI.

¹⁴²⁶ Above n 1424.

¹⁴²⁷ Above n 1424.

¹⁴²⁸ Ronald Arkin of the Georgia Institute of Technology has argued that AI-powered military robots might in fact be ethically superior to human soldiers; they would not rap, pillage or make poor judgments under stress.

¹⁴²⁹ An option suggested in the civil context by the Draft Report of the European Parliament's Committee on legal Affairs with recommendations to the Commission Civil Law Rules on Robotics 2015/2103(INL), para [27].

in which case we either have to accept decriminalisation altogether or we run the risk of blurring the effective signalling function performed by criminal law in its core cases which are 'truly criminal'. Or, if the offence remains 'truly criminal' a move to strict liability is highly problematic. Mens rea is central to the criminal law's entitlement to censure¹⁴³⁰ and we cannot simply abandon that key requirement of criminal liability in the face of difficulty in proving it.

Nor would such a move necessarily be sufficient, since even offences of strict liability require proof of some kind of causation or at least control by the defendant. Thus, for example in *R v Robinson-Pierre*¹⁴³¹ the defendant was charged with an offence contrary to s 3 of the 1991 Dangerous Dogs Act. This provides that

If a dog is dangerously out of control in [any place in England or Wales (whether or not a public place)]

—

- (a) the owner; and
- (b) if different, the person for the time being in charge of the dog, is guilty of an offence, or, if the dog while so out of control injures any person [or assistance dog], an aggravated offence, under this subsection.

So there is no reference in this section to any requirement for the defendant to have caused the dog to be out of control, nor for the defendant to be aware of that fact. Nonetheless, in *Robinson-Pierre* the police forcibly entered the defendant's home, where they were attacked by his pit-bull terrier which followed them onto the street and on appeal the Court held that in these circumstances the defendant did not have sufficient control over the situation to be liable for the offence.

There is one exception to this approach. In the environmental context, in *Empress Car*¹⁴³² the House of Lords held that the defendants were liable for an escape of oil into a river, even though the escape of the oil was actually caused by the vandalism of a third party. This is generally regarded by criminal law scholarship¹⁴³³ as an aberrant and incorrectly decided case in the law of causation, but it may point to a potential solution in cases of ML decision-making.

The approach generally preferred by criminal law scholarship to that of strict liability is one of negligence, or lack of due diligence.¹⁴³⁴ This avoids the possibility that those who are wholly innocent will be convicted, while imposing less of a burden on the prosecution than a full finding of intent or knowledge

¹⁴³⁰ See in particular A Ashworth, 'Should Strict Criminal Liability be Removed from All Imprisonable Offences?' (A Ashworth, *Positive Obligations in Criminal Law* (2013) 121.

¹⁴³¹ [2013] EWCA Crim 2396.

¹⁴³² *Environmental Agency (formerly National Rivers Authority) v Empress Car Co (Abertillery) Ltd* [1999] 2 AC 22.

¹⁴³³ See e.g. D Ormerod and K Laird, *Smith and Hogan's Criminal Law* (2015) pp 99-100.

¹⁴³⁴ See, e.g. Ashworth, above n 1430, Ormerod and Laird, *ibid* at 199-203.

might do. A good example of this is s 7 of the Bribery Act 2010 which creates a strict liability offence for a commercial organisation where a person associated with it bribes another person intending to obtain or retain a business advantage. The offence itself is very wide, but s 7(2) provides that it is a defence for the commercial organisation to prove that it had in place adequate procedures designed to prevent persons associated with it from undertaking such conduct. This does provide one possible solution to the problem of ML regulation infringement, but there are two potential dangers associated with it. One is that even if the defendant puts in place adequate procedures, while these might be sufficient to prevent wrongdoing in the case of human beings, who are capable of understanding the spirit of a regulation as well as its letter, the same may not be true of ML systems. There may therefore be a mismatch between the steps it is reasonable to expect a corporation to take and the means which would actually be necessary to prevent a significant amount of infraction by ML systems. Another is that in some instances, such as the Food Safety Act 1990, strict liability offences with reasonable precaution defences are accompanied by the rule that defendants may not rely on the reasonable precaution defence when the allegation is that the offence resulted from the actions or inactions of a third party unless within a prescribed period the defendant has served on the prosecutor a notice in writing identifying that actual wrongdoer. If applied to those programming the system it carries the dangerous possibility that those people will be used as scapegoats for a form of infraction which benefited the company as a whole, but the chances are that even in such a context it would be difficult to ascertain which programmer was responsible for a particular line of code, or indeed the extent to which the resulting programme was the result of the initial code or the subsequent development of that code by the ML system.¹⁴³⁵ And of course if the identification requirement were applied instead to the ML system it would be pointless.

In a sense, however, both these negligence/lack of due diligence offences and the approach taken in the environmental context in the case of *Empress Car* may point towards another, more helpful suggestion for responding to wrongdoing by ML systems. What is being targeted in both these approaches is arguably a *failure* on the part of the relevant managers to take sufficient precautions. Arguing for such a standard is therefore akin to arguing for a move towards greater liability for omissions. This too is a historically controversial aspect of criminal law, but it is arguable that what Ashworth calls 'conditional positive obligations' are perhaps less controversial than some other examples: 'if a person undertakes a certain activity or enters a certain business, he or she should expect to take on certain duties'.¹⁴³⁶ Particularly, one might suggest, when the person or persons in question derive a significant benefit from the

¹⁴³⁵ Not least because it is possible for ML systems to write their own code, piecing together lines of code taken from existing software. <https://www.newscientist.com/article/mg23331144-500-ai-learns-to-write-its-own-code-by-stealing-from-other-programs/>

¹⁴³⁶ Ashworth, above n 1430 at 79.

activity or business which creates the risk. However, here too the autonomy of ML may present us with a problem in terms of drafting the relevant controls. One option for omissions offences, and that which is closest to the 'due diligence' approach outlined above, is to criminalise the omission *itself*. The other option is to render the defendant liable for omitting to prevent the relevant harm. But as Ashworth points out, both of these approaches ought in principle to be subject to a defence of impossibility, in the sense that a defendant should not be held liable for something he, she or it could not have prevented, and it seems very likely that that impossibility might arise in the event of ML decision-making. And on the contrary if we focus simply on the omission itself, e.g. an omission to take reasonable care, then we have the same problem as outlined above in relation to due diligence defences which are the functional equivalent; however much care or due diligence the defendant exercises, it may not be nearly enough in many cases, and the danger is that the whole process would resort to a box-ticking exercise by the relevant company, rather than any real attempt to prevent harm or damage.

A more promising alternative may therefore be to focus omissions liability not on the period immediately preceding the wrongdoing, but that immediately following it. Building on Fisse and Braithwaite's concept of 'reactive fault',¹⁴³⁷ one option for future regulation might be to examine whether companies do monitor their own ML systems to establish whether they are acting in accordance with the law. Rather than requiring firms to demonstrate an almost impossible level of due diligence aimed at preventing the wrongdoing in the first place, or accepting a relatively meaningless formal approach to ex ante due diligence, it would be possible to impose a much more stringent kind of ex post due diligence. How rapidly did a car manufacturer detect that its cars were breaching emissions rules? How rapidly did a self-driving car hire service pick up on the fact that its cars were evading speed limits and engaging in other kinds of dangerous driving? How rapidly did a group of companies pick up on the fact that their AI systems were co-ordinating in a manner likely to breach antitrust or competition rules? Failure to detect and respond swiftly to such failures seems likely to provide a solution which is both practically more feasible and therefore inherently more likely to address the true wrongdoing for which it is appropriate to hold the individual manager or corporation liable. And that in turn seems likely to preserve the core integrity and thus the key signalling function of criminal law.

8 September 2017

¹⁴³⁷ B Fisse, 'Reconstructing Corporate Criminal Law: Deterrence, Retribution, Fault and Sanctions' (1983) 56 *Southern California Law Review* 1141; J Braithwaite, 'Intention Versus Reactive Fault', Ch 16 of *Intention in Law and Philosophy*, (ed N Naffine, R Owens and J Williams) Ashgate, Dartmouth, 2001.

Professor Michael Wooldridge – Written evidence (AIC0174)

I am pleased to hereby provide a written submission to the House of Lords Select Committee on Artificial Intelligence, in response to your Call for Evidence of July 2017. Below I respond to the questions raised in turn; at the end of the document I provide a brief statement regarding my credentials for this task. My text is in **bold face**.

1. *What is the current state of artificial intelligence and what factors have contributed to this? How is it likely to develop over the next 5, 10 and 20 years? What factors, technical or societal, will accelerate or hinder this development?*

To answer this question we need to distinguish between “General Artificial Intelligence” (AI) and “Narrow AI”. General AI is the long-term dream of AI researchers: to build machines that are conscious, self-aware – machines that have “general” intelligence, in broadly the same way that people do. In contrast, Narrow AI is about getting machines to solve *specific tasks which currently require brains*. Examples of such tasks might be recognising faces, driving cars, automatically dictating spoken text, and translating texts from one language to another.

There has been real, and dramatic progress in Narrow AI since the turn of the century. The current state of the art in automated translation, for example, would have astonished researchers just two decades ago.

However, there has been essentially no progress in General AI, and I see no likelihood of such in the immediate future.

I believe there will continue to be progress in Narrow AI, and we will see AI techniques embedded ever more widely. We will not, in the next decade, see anything like human-level General AI

There are 3 key drivers behind recent advances: (i) a string of scientific developments, which made it possible to apply “machine learning” techniques to more complex problems than had hitherto been considered possible; (ii) the availability of “big data”, which is required to “train” AI systems; and (iii) the availability of cheap computer processing power, required to train AI systems using data.

Over the next decade, all these trends will continue, and we will see AI applied to ever more complex problems, which will continue to make headlines.

2. *Is the current level of excitement which surrounds artificial intelligence warranted?*

We should be excited about AI in the same way we were excited about the arrival of microprocessors in the 1970s, desktop computers in the 1980s, the Internet in the 1990s, and smart phones and mobile computing in the early part of this century. Like all of these technologies, AI is *generic*, in the sense that it has a very wide range of possible applications – many of which we can't foresee at present. These applications will improve our lives, create new economic opportunities, and lead to wealth creation – but will surely also lead to redundancies and businesses closing down.

However, while I believe there is genuine cause for excitement, it is clear that there is an AI bubble at present, with unrealistic expectations particularly about "General AI" (see above). There have been many uninformed and unwarranted comments about AI in the press over the past 3 years, which I very much regret.

So, while we should be excited about this new technology, we should not lose sight of the fact that this is all it is: a new technology.

3. How can the general public best be prepared for more widespread use of artificial intelligence?

The best guard against the economic challenges raised by AI is simply education. Automation will continue to take jobs for the foreseeable future. The best insurance against redundancy due to automation is education.

4. *Who in society is gaining the most from the development and use of artificial intelligence and data? Who is gaining the least? How can potential disparities be mitigated?*

At present, those who are gaining the most are global data-rich companies like Google, facebook, and the like. They gain competitive advantage by using AI to offer services that their competitors cannot.

The use of AI based on big data makes it harder for competitors to enter the marketplace. While startups developing niche AI applications (such as SwiftKey) can prosper, it becomes increasingly hard to challenge the global dominance of large companies – the best they can hope for is to be acquired or sell services to the giants.

5. *Should efforts be made to improve the public's understanding of, and engagement with, artificial intelligence? If so, how?*

I believe it is very important to responsibly inform the public of the reality of AI today – what it can do, what it cannot, what we should worry about, and what we should not.

There are excellent precedents in the UK: the BBC's micro initiative in the early 1980s prepared a generation for the desktop computer revolution, and surely paid for itself many times over in the long run.

6. *What are the key sectors that stand to benefit from the development and use of artificial intelligence? Which sectors do not?*

The most immediate and rewarding benefits lie in healthcare – at individual and national levels.

At the level of individuals, smartphones and smart appliances (such as the Apple Watch) can provide healthcare apps with a stream of data about us, and AI technologies can interpret this to provide real-time healthcare advice. Personalised smart healthcare apps will take the "fitbit" experience into an entirely different league. It is entirely plausible that we will have apps that can predict the onset of dementia, for example, simply based on changes in the way that we interact with them; or that we will have cars that can refuse to allow us to drive them because they judge we are not in a fit state to do so. However, the insurance industry can surely also see the potential uses of real-time data streams about us, some of which we might judge to be inappropriate. Government needs to consider and potentially legislate over these issues.

In terms of national healthcare, the NHS is uniquely positioned as a national healthcare provider to make use of AI techniques for example by automating some services, and by augmenting doctors and nurses in others. For example, one of the biggest and most expensive bottlenecks in the use of data provided by X-rays the like is the time required by highly trained staff to interpret the

data. It is entirely realistic that we can augment physicians with software that assists in such tasks.

Beyond this, though, I think it is important to emphasise again that AI is a *generic* technology. It can be applied anywhere that decisions have to be made, based on experience and judgement.

7. *How can the data-based monopolies of some large corporations, and the 'winner-takes-all' economies associated with them, be addressed? How can data be managed and safeguarded to ensure it contributes to the public good and a well-functioning economy?*

I see no obvious solution to the hegemony of the global data giants. All we can do is ensure that the rights of individuals and organisations with respect to their data are upheld and protected.

In terms of protecting individuals and organisations, the UK already has excellent legislation in the form of the Data Protection Act 1998. This covers most of the main issues, and remains surprisingly robust, given the many technological changes since it was introduced.

However, this requires resources, a pro-active approach by regulators, and the political will to challenge bad practice even where this is seen as economically or politically awkward. If we are prepared to hold companies to account against the existing legal framework, this would address many of the concerns that we all have about our data.

One specific difficulty that is increasingly a concern is that data doesn't respect national borders. Most users of DropBox, for example, would be hard-pressed to tell you where the servers are that store their data, or which nation's legislation governs the storage of it. International cooperation is the only answer here, I believe.

8. *What are the ethical implications of the development and use of artificial intelligence? How can any negative implications be resolved?*

One obvious issue is in lethal autonomous weapons – that is, weapons that make the decision whether to take human life. This issue needs proper public debate – the widespread development of autonomous weapons by rogue states would be as worrying as the development of chemical, biological, and nuclear weapons by such states. Political pressure to prevent the proliferation of such weapons might therefore be appropriate – in the same way that

political pressure is applied to prevent the development of chemical, biological, or nuclear weapons.

Beyond this, most of the issues seem to be further developments of existing issues, such as the ownership and protection of data relating to private individuals, as discussed above.

9. *In what situations is a relative lack of transparency in artificial intelligence systems (so-called 'black boxing') acceptable? When should it not be permissible?*

Lack of transparency is an issue wherever the decisions made or actions taken by AI systems will have substantive real-world consequences for individuals or organisations.

The issue here simply seems to be that those who use AI (or any other algorithmic solution) to make such decisions have to clearly present the criteria by which the decisions will be made, and must be ready to defend them. Legislation in areas such as insurance, healthcare, employment and so on may be necessary to support this.

10. *What role should the Government take in the development and use of artificial intelligence in the United Kingdom? Should artificial intelligence be regulated? If so, how?*

The UK is in the enviable position of being a world-leader in an exciting new technology. While we have often been in such a position with new technologies (including computers), we have rarely managed to maintain it. I think that it is therefore essential that the Government nurtures this area. Specifically: (i) invest in R&D around areas where the UK has genuine potential, and build up capacity; and (ii) do everything possible to nurture the UK's burgeoning start-up sector (one issue here is Brexit – this could quite genuinely be the death knell for UK tech startups, which are *heavily* reliant on overseas talent).

I don't see regulation is necessary beyond the maintenance of our excellent Data Protection Legislation, and potentially a few specific areas as described above. It is important to avoid knee-jerk reactive legislation such as "robot taxes", as I believe these will not ultimately be in the UK's best interests.

11. *What lessons can be learnt from other countries or international organisations (e.g. the European Union, the World Economic Forum) in their policy approach to artificial intelligence?*

The UK is leading the way in AI, and we have excellent data protection legislation. I don't see specific initiatives to copy from elsewhere, although I do think international cooperation (with respect to issues discussed above) will be required to handle issues such as global data. (While I am a Europhile, I think some of the suggestions coming from the EU with respect to AI are naïve and unworkable.)

ABOUT ME. I am a Professor of Computer Science and the Head of Department of Computer Science at the University of Oxford. I have been an AI researcher for 25 years, and have published about 400 scientific articles on this topic; I am one of the UK's most cited researchers in AI. I was President of the European Association for AI (EurAI) from 2014-16, and President of the International Joint Conference on AI (IJCAI) from 2015-17. I am a Fellow of the Association for Advancement of AI (AAA), the European Association for AI, and the Association for Computing Machinery (ACM).

Michael Wooldridge
Professor and Head of Department of Computer Science
University of Oxford

6 September 2017

Workday Inc. – Written evidence (AIC0183)

1. Workday, Inc., a publicly traded, global enterprise cloud application provider for human resources and finance, welcomes the opportunity to provide feedback to the Select Committee on Artificial Intelligence in response to the July 19, 2017, Call for Evidence.
2. Workday supports the Committee’s efforts to obtain feedback from industry, and believes a collection of evidence from all sectors can help shape a more effective discussion on Artificial Intelligence (“AI”). Given rapid technological advancements and the potential transformative effects of AI, excitement in the area is warranted. Because of these rapid advances, sure to accelerate over the next two decades, legislating in this area presents challenges: concerns over potential unintended consequences of AI must be weighed along with the many socioeconomic benefits AI can offer. To maintain pace with technology, Workday supports the development of an industry-driven Code of Conduct.
3. While the Committee poses several questions with respect to AI, our comments focus on (i) clarifying the distinctions between two often-conflated concepts, big data analytics and AI; and (ii) articulating why an industry-driven Code of Conduct focused on the ethical and responsible development of AI, rather than formal regulation, is the best way forward in this area.

The Distinction Between Big Data Analytics and Artificial Intelligence

4. Big data analytics and AI, while two separate but related areas of computer science, are often used interchangeably in policy discussions. This trend is exacerbated by the fact that even sources that define these terms differ from one another on their precise meanings. Workday believes that clarifying the distinctions between big data analytics and AI will help in identifying the potential benefits each area offers, as well as the potential risks – thereby better informing both the public’s understanding of these technologies and discussions of their policy implications.
5. Specifically, big data analytics refers to analysing and leveraging large quantities of data to provide instantaneous, accurate, and useful insights, create correlations, and allow individuals and businesses to make better decisions.¹⁴³⁸ For example, with Workday’s Data-as-a-Service (“DaaS”), companies can contribute data to a collective community and then benchmark themselves against other participants to gain helpful insight into overall

¹⁴³⁸ Adapted definition from Merriam-Webster, which defines “analytics”. See Merriam-Webster, Analytics, available at <https://www.merriam-webster.com/dictionary/analytics> (last visited 6 September 2017).

performance, growth, retention rates, as well as access other useful metrics and trends.¹⁴³⁹ Another example of big data analytics is Workday’s Retention Risk Analysis, which uncovers trends in historical data about turnover in order to yield predictive insights that help decision-makers analyse, answer, and act on questions.¹⁴⁴⁰

6. This last point is critical. Big data analytics provides insights that were previously unavailable, allowing people to make better decisions. Even where machine learning is used, the end result is the provision of real-time, drillable information that is actionable by human beings.
7. By contrast, AI is “a [separate] branch of computer science dealing with the simulation of intelligent behaviour in computers”¹⁴⁴¹ and “[t]he theory and development of computer systems able to perform tasks normally requiring human intelligence.”¹⁴⁴² AI may also be defined as “the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence”.¹⁴⁴³ Some examples of AI include self-driving cars, intelligent translation tools,¹⁴⁴⁴ and medical “chatbots”.¹⁴⁴⁵ In turn these implementations of AI benefit consumers, whether through increased safety,

¹⁴³⁹ Workday Unveils Data-as-a-Service Based on Customer Demand, *available at* <http://blogs.workday.com/workdays-unveils-data-service-based-customer-demand/> (last visited 6 September 2017).

¹⁴⁴⁰ Workday Delivers First Wave of Insight Applications; Professional Services Suite, *available at* <https://www.workday.com/en-us/company/newsroom/press-releases/press-release-details.html?id=1940591> (last visited 6 September 2017).

¹⁴⁴¹ Merriam-Webster, Definition of Artificial Intelligence, *available at* <https://www.merriam-webster.com/dictionary/artificial%20intelligence> (last visited 6 September 2017).

¹⁴⁴² Oxford University Press, Oxford Living Dictionaries (2017), *available at* https://en.oxforddictionaries.com/definition/artificial_intelligence (last visited 6 September 2017).

¹⁴⁴³ John McCarthy, What is Artificial Intelligence?, Stanford University, Computer Science Department, November 12, 2007, *available at* <http://www-formal.stanford.edu/jmc/whatisai/node1.html> (last visited 6 September 2017). One of the most effective ways to determine whether machine intelligence exists is by applying the Turing test, created by Alan Turing, arguably the father of computer science and AI. The Turing test is “a proposed test of a computer’s ability to think, requiring that the covert substitution of the computer for one of the participants in a keyboard and screen dialogue should be undetectable by the remaining human participant to determine whether the work of a computer constitutes AI”. See *generally*, A.M. Turing, Computing Machinery and Intelligence (1950), *available at* <http://loebner.net/Prizet/TuringArticle.html> (last visited 6 September 2017).

¹⁴⁴⁴ See *e.g.*, The Next Web, Facebook Researchers use AI to Build a Better Translator, August 3, 2017, *available at* https://thenextweb.com/artificial-intelligence/2017/08/04/facebook-turns-translation-duties-over-to-the-machines/#.tnw_4bKqNAPy (last visited 6 September 2017).

¹⁴⁴⁵ See *e.g.*, Forbes, See How Artificial Intelligence Can Improve Medical Diagnosis And Healthcare, May 16, 2017, *available at* <https://www.forbes.com/sites/jenniferhicks/2017/05/16/see-how-artificial-intelligence-can-improve-medical-diagnosis-and-healthcare/> (last visited 6 September 2017).

better communication, or more timely diagnoses, to take the examples above.

8. While AI relies on big data to power its functionality – after all, without data AI would have nothing on which to exercise its “intelligence” – it differs from big data analytics in that the decisions are automated. Whether executing turns by driving, adjusting thermostats, or determining when to water crops, AI acts for humans (albeit under their ultimate control). Use of AI poses a specific set of policy questions related to automated decision-making, which has implications ranging from data privacy, where automated decision-making regarding people is already covered by the General Data Protection Regulation, to workforce development, where there are concerns about job losses resulting from increased deployment of AI. By contrast, with big data analytics the key questions relate to the quality and availability of data – for example, where do data sets originate and whether they are free from bias.
9. Given these distinct questions, applying a regulatory regime designed for AI to big data analytics would be a poor fit and could retard the growth of big data analytics and the benefits that all of us gain from new insights. As the House of Lords considers these issues, it should carefully distinguish AI and its policy implications from those related to data analytics.

An Industry-Driven Code of Conduct

10. Although the concept of AI has arguably existed since the 1950s,¹⁴⁴⁶ we remain in the early phases of its technology and development. Today, we simply lack full knowledge and understanding of the field and its potential as well as its implications. As such, regulating AI at this juncture could prove ineffective at best and has the real potential to hinder innovation in ways unknown. Moreover, a variety of existing laws, regulations, and policies in areas such as business practices, data protection, and cybersecurity already address many concerns raised by AI.
11. That said, it is also clear that development of AI needs to occur thoughtfully, and account for the legitimate concerns raised by various observers. In our view, the best way to address this need while remaining flexible and nimble as technology continues to evolve is through development of an industry-driven Code of Conduct. In creating a Code of Conduct, government and industry should seek input across all sectors of the economy and society, to ensure thoroughness and representation of all viewpoints. AI’s impact extends beyond the technology industry, and any effective Code must likewise reflect the diverse interests of society.

¹⁴⁴⁶ See generally, A.M. Turing, *Computing Machinery and Intelligence* (1950), available at <http://loebner.net/Prizef/TuringArticle.html> (last visited 6 September 2017).

12. In addition, as an essential foundation of an AI Code of Conduct, industry must focus on ethical and responsible development. Ethical development of AI embeds certain basic principles into the foundation of the computer's intelligent decision-making. These principles include basic human rights, gender and ethnic equality and minimizing bias, protection of individual rights and freedoms, protecting against unintended negative consequences, maintaining human control of AI, avoiding reasonably predictable misuse of AI, and ensuring human safety and full context is understood prior to any automated decision making.¹⁴⁴⁷ These principles should be embedded by design throughout the lifecycle of any AI implementation.
13. Finally, in creating a Code of Conduct, industry must also recognize its obligation to remain compliant with all existing applicable laws, regulations, and policies, including those related to business operations as well as global data protection and cyber-security laws (*e.g.*, the General Data Protection Regulation).
14. AI offers great potential for the betterment of society, the furthering of human progress, solutions to some of our most important challenges, and opportunities from which both individuals and businesses can benefit. To ensure society and the workforce develops relevant, meaningful skills and is prepared for AI, governments should maintain an open dialogue with industry, continue to invest in AI research, and ensure innovation in the area continues. In addition, as noted above, there must be a clear and distinct understanding of the differences between big data analytics and AI.
15. With the many benefits of AI, regulating too soon in this area may prove ineffective, hinder innovation, and prevent positive developments in the field. Given the early state of AI technology, global governments should, in continuing engagement with industry and society, support an industry-driven Code of Conduct, focused on ethical and responsible development.
16. Workday appreciates the opportunity to share our views with the House of Lords in response to this consultation, and we would welcome the opportunity to continue the discussion of the policy implications both big data analytics and AI with the House of Lords and other branches of the UK Government.

6 September 2017

¹⁴⁴⁷ See also World Economic Forum, Top 9 Ethical Issues in Artificial Intelligence (October 21, 2016), available at <https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/> (last visited 6 September 2017).

Young Enterprise – Written evidence (AIC0091)

Introduction to Young Enterprise’s work on artificial intelligence

1. Young Enterprise welcomes the opportunity to respond to the call for evidence by the House of Lords Select Committee on Artificial Intelligence.

2. At Young Enterprise we work directly with young people (ages 4 to 25) and their teachers, in schools across the UK. We have been helping young people to transition from education to employment for over 50 years and recognise the importance of ensuring that students are aware of the realities of a changing world of work, as well as helping them to develop the crucial employability skills that will help them to secure work and be successful in it. At Young Enterprise, we have identified eight core employability skills, essential for unlocking young people’s potential and enabling them to thrive throughout their careers: communication, confidence, financial capability, initiative, organisation, problem solving, resilience and teamwork.

3. We’re conscious of the potential for artificial intelligence (AI) to transform the workplace. In March 2017 research from PwC¹⁴⁴⁸ revealed that approximately 30% of UK jobs will be susceptible to automation from robotics and AI by the early 2030s, with automation most likely in sectors such as transport, manufacturing and retail. A separate report from Reform¹⁴⁴⁹, published in February 2017, suggested that the public sector could lose around 250,000 jobs in admin functions by replacing human beings with robots.

4. In July 2017, we polled young people to find out what they thought about AI and the robot revolution. The report, entitled Robot Revolution: The impact of artificial intelligence on entrepreneurs and job prospects¹⁴⁵⁰, includes detailed polling of 200 finalists from the Young Enterprise flagship ‘Company Programme’, from all regions of the UK, shortlisted from 20,000 entrants. Some of the key findings were:

- When asked about the impact of AI in the workplace, 76 per cent said they believed fewer jobs would be available due to the use of robots in the workplace. In contrast just 10 per cent said this trend would lead to more jobs and 14 per cent said there would be no noticeable change.
- When asked about the role of robots in the workplace, nearly half of respondents (47 per cent) said they were ‘concerned’ at the prospect of machines occupying a large percentage of the workforce. 35 per cent said they were neutral, and only 18 per cent felt comfortable.

¹⁴⁴⁸ PwC, UK Economic Outlook, March 2017

¹⁴⁴⁹ Reform, Work in progress. Towards a leaner, smarter public-sector workforce, February 2017

¹⁴⁵⁰ Young Enterprise Robot Revolution: The impact of artificial intelligence on entrepreneurs and job prospects, July 2017

- 59 per cent of respondents said that they thought it would be harder to get a job that a robot could also do, due to a lack of basic core skills like team work and problem solving.
- When asked if they would accept a job working for a robot 45 per cent said yes whereas 55 per cent said no.

5. Our Young Enterprise programmes are designed to prepare young people for the world of work, developing knowledge of the workplace and the key employability skills required to succeed in and adapt to a changing workplace. Our key initiatives include:

- **Company Programme** – our flagship programme which allows 15-19 year olds to set up their own business for a year and make all the decisions about the company, from raising the initial share capital through to designing their own product, selling directly to customers and paying their taxes. Over the course of the 2016/17 academic year, 20,000 students from around 2,000 student teams across the UK set up and ran a student Company with Company Programme. Many student companies develop digital or technology innovative businesses including mobile apps, virtual reality headsets or innovative uses for contactless technologies.
- **Team Programme** – a tailored enterprise scheme allowing young people with learning difficulties or disabilities to set up their own business.
- **Fiver** and **Tenner** – nationwide competitions where young people at primary and secondary levels create a new business with just £5 or £10 and compete to make the most profit and social impact over the course of a month.
- **Learn to Earn** – one of our most popular one-day programmes which helps students to think about their career choices and what kind of qualifications they would need to pursue them, including how apprenticeships may help them to achieve their ambitions. Students are guided to learn more about themselves, consider their own strengths and build greater self-awareness, and to explore career options aligned to personality traits.

6. Young Enterprise believes that there is a need for schools to place a greater emphasis, from an early age, on preparing young people for their future careers. We believe that through engagement with employers and programmes designed to develop employability skills we can better prepare young people for a world of work where AI use is widespread.

7. Our recommendations for helping young people succeed in an AI dominated workplace are outlined in further details below, primarily aimed at question (3) of the Call for Evidence: 'How can the general public best be prepared for more widespread use of artificial intelligence?'

Importance of increasing raining for pupils in soft skills

8. Increasing capabilities of AI look set to transform many traditional jobs beyond all recognition in addition to creating unforeseen new jobs. Young people today will be expected to reinvent themselves multiple times throughout their careers as artificial intelligence replaces existing jobs and creates new jobs.

9. At Young Enterprise we know that 'learning by doing' activities have a significant positive impact on young people's employability skills, ambitions and outlook on the world of work. Our impact analysis of our 2015/16 Company Programme found that:

- 95% of participants agreed that they had improved at least one employability competency including self-esteem, aspirations and work readiness.
- 74% of teachers agreed that the programme had contributed to raising students' career aspirations.
- 65% of participants agreed that their involvement in the programme had helped them feel more ready for the world of work.

10. The tangible benefits of developing these skills are significant - 94% of Young Enterprise Company Programme Alumni are in an Education, Employment or Training destination - 7% higher than the national average.

11. Employers are already demanding employability skills and flexibility, in July 2017 a CBI/Pearson Education and Skills survey¹⁴⁵¹ found that employers demand skills such as resilience, confidence and communication and rate attitude and aptitude for work as more important than academic achievement when recruiting school and college leavers. These skills, which will enable young people to stand out when competing for jobs with machines, will likely become increasingly necessary as AI in the workplace becomes widespread.

Importance of increasing employer engagement in schools

12. Young Enterprise, through our work in schools across the country, has found engagement with a range of inspiring mentors from the world of work can inspire young people to discover new careers, and empower them to raise their aspirations and unlock their potential.

13. We work with over 7,000 business volunteers who provide students of all ages with a diverse mix of role-models from a range of different backgrounds, many of them from a technological professional background. These volunteers are able to draw upon their experience of working in a variety of sectors and

¹⁴⁵¹ CBI/Pearson Education and Skills survey, July 2017

professions to help inspire the next generation. It is vitally important that businesses with an interest in AI are involved in such initiatives.

14. The benefits of employer engagement has been researched by Education and Employers¹⁴⁵² who, in January 2017, found students who undertake at least three school-mediated employer engagements are 85% less likely be NEET. We believe that organisations, especially businesses with an interest in AI, should be encouraged to enter schools via enterprise initiatives to help inspire and prepare young people for the future world of work.

Policy recommendations

15. Young Enterprise strongly believes, particularly as AI in the workplace becomes increasingly widespread, that young people need to start learning about the world of work and developing employability skills from an early age. These skills take several years to develop, therefore students from an early age in schools should be provided with the opportunity to develop these 'soft' skills. In the interest of UK productivity as well as social mobility it should not be left to fluke of postcode lottery or family background as to whether a young person is provided with the opportunity to grow these crucial skills. The policy proposals below provide practical opportunities to improve careers and skills education in schools:

- i. Young Enterprise welcomed the government's announcement to look at providing statutory status for PSHE with a strong 'E' for economic strand ensuring young people would be taught economic wellbeing, financial capability and careers preparation . By providing statutory status for PSHE and ensuing the 'E' strand – for economic – lies at the heart of the subject, the government can help ensure all young people develop the skills, knowledge and confidence they need to succeed in a changing world of work.
- ii. A longer term focus for destinations data should be adopted, including greater monitoring of young people's satisfaction within their chosen career path. Destinations data is currently limited to looking at young people's destinations one year after finishing Key Stage 4 or Key Stage Five. In contrast the Government's Longitudinal Educational Outcomes (LEO) study looks at university graduates one, three and ten years after graduation.
- iii. Schools should be required to publicise both their approach to preparing young people for the world of work and their longer-term destinations data to help inform parents in their school selection. Schools would consequently be encouraged to increase employer engagement, develop

¹⁴⁵² Education and Employers Contemporary transitions: Young Britons reflect on life after secondary school and college, January 2017

soft skills and prepare young people for the challenges of a workplace with widespread AI use.

- iv. School should be encouraged to appoint a lead teacher on the 'Senior Leadership Team' for financial and enterprise education to coordinate provision in this area, as well as a lead Governor.

16. Young Enterprise hopes this response is of assistance to the Select Committee on Artificial Intelligence and would be happy to expand on any of the issues raised above if that were of interest.

5 September 2017

Dr Jianhan Zhu – Written evidence (AIC0045)

Will AI systems keep growing in popularity?

Artificial intelligence (AI) and machine learning has attracted lots of attentions in recent years due to impressive applications and results in scientific research, and industrial and social areas. An increasing number of countries are putting more and more emphasis and investments in AI, hoping to take advantageous positions in future AI technologies. Many deem AI's development as the next industrial revolution, and it could change the society completely. We have only seen the beginning of the changes that AI will bring to us, and its developments in the next 5 to 20 years will be at an even more accelerated pace. Given the dilemma of immaturity of current AI technologies and their great potential, we are still faced with many challenges in terms of ethics, security, and privacy etc. The UK has so far established a leading position in AI research and applications, and it is worthwhile to sustain and increase the leadership by investing more in this important area.

Potential of AI systems

Artificial intelligence is not a newly coined term, and the idea traces back to the 13th-century Ramon Llull. AI algorithms such as neural networks proposed before have become more popular due to much increased computational power in recent years. Although the basics of AI are much simpler than human's intelligence (neural networks were inspired by human brain but are nothing near brain's complexity), powerful computation can still make basic algorithms to do extraordinary things (such as AlphaGo's records of beating top human players). With the ever-increasing computational power, it is possible that current AlphaGo running on thousands of cloud computers could instead be running on a handheld device in future. This together with more advanced AI algorithms (e.g., more complex neural networks) and more data available to train these algorithms (given the increasing popularity of sensors and social networks which collect data), gives a picture of more powerful and ubiquitous AI systems in future.

Understanding how AI systems work

Efforts so far have focused more on developing AI systems than with understanding how they work. For example, AlphaGo could make moves that humans could not understand. There have been efforts to interpret the reasons behind AI systems' decisions (e.g., Nvidia self-driving cars visualise determining road features in their decisions and compare the features with humans'). It would be fine to have black boxes in games. But for real world applications such as medical and military applications, technologies to understand reasoning mechanisms of AI systems are essential for ethical and legal reasons.

Technologies to prevent AI systems' bad behaviours and monitor AI systems to keep them in control are also important. Failure to develop above mentioned technologies can hinder their adoptions and applications.

Companies such as Google have established ethics committees to look at impact of AI technologies. Understanding of how AI systems work paves the way to make them ethical and accountable.

Involving humans in the decision-making process by AI systems is important for safety. Humans should be helped to understand or have assurances of the decisions made by the systems. The systems by design should help humans understand their decisions.

Enhancement of human intelligence

AlphaGo initially trained its neural networks based on previous human players' games. But it then went on to train by playing against another AlphaGo player to improve. Machines are progressing to gain new knowledge by training with themselves. Top human players have also improved their skills by learning from AlphaGo. It could be beneficial for machines to learn by themselves, and what they learn can add to existing human knowledge and help humans to improve.

Machines could find new knowledge missed by humans, and machines can physically get to places unsuitable for humans (such as the Mars or nuclear reactors) and learn in these environments. Machines can act on behalf of humans, and extend human's intelligence and capabilities.

Elon Musk argued that humans need to merge with machines to become relevant, otherwise machines might take on their own evolutionary paths and leave humans behind. Recently, he founded the Neuralink to create an interface between the human brain and machines to firstly venture into helping people with medical conditions, and then go on to merge human and machine intelligence. There will be technological challenges as well as ethical and legal issues related to this science fiction becoming reality.

Disruptive nature

Few predicted that AI systems could beat top human Go players so soon. It is very hard to predict how AI systems will evolve, and it is very possible that new AI technologies will surprise us all. It is important to pay attention to developments in AI in academia and industries, and to have a mechanism to respond and adjust quickly to new developments in AI.

How to define artificial intelligence?

AI is concerned with human made algorithms, mechanisms and systems that can exhibit intelligent behaviours. AI systems have been built by humans to help them complete meaningful tasks. An important role of AI systems is to be helpful and beneficial to humans. An important factor in AI system development is to maximize their benefits and at the same time minimize their negative effects. AI systems can extend and enhance human intelligence. They originate from humans, but they could help humans improve themselves, and understand themselves and the nature.

Conclusions

Development of AI has provided unprecedented opportunities and challenges. Stephen Hawking warned that advanced AI would be “either the best, or the worst thing, ever to happen to humanity”. We are now in the position to determine what the future of AI will be, and how humans and machines will coexist. With the right approach, monitoring and safety measures, we have every reason to believe that AI technologies will benefit humans, and help humans build better lives and societies.

Biography

Dr Jianhan Zhu has a PhD in Computer Science. He has many years of experience working in information retrieval, data mining, machine learning, artificial intelligence, social network, and semantic web related technologies in both academia and industries. He has published over 40 papers in international journals and conferences.

1 September 2017

Diego Zuluaga – Written evidence (AIC0235)

The economic and policy consequences of AI - written submission

Diego Zuluaga, Head of Financial Services & Tech Policy, Institute of Economic Affairs

About the author

Diego Zuluaga was educated at McGill University and Keble College, Oxford, from which he holds degrees in economics and finance. His policy interests are mainly in consumer finance and banking, capital markets regulation, and multi-sided markets. However, he has written on a range of economic issues including the taxation of capital income, the regulation of online platforms and the reform of electricity markets after Brexit. Diego's articles have featured in UK and foreign outlets such as Newsweek, City AM, CapX and L'Opinion. He is also a frequent speaker on broadcast media and at public events, as well as a lecturer at the University of Buckingham.

DISCLAIMER: As part of its educational objectives the IEA facilitates responses to public policy consultations by academics and others. However, the views expressed, whilst generally consistent with the IEA's mission, are those of the authors and not those of the IEA (which has no corporate view), its managing Trustees, senior staff or Academic Advisory Council. If these views are quoted then we ask they are quoted as the views of the author(s)

Does artificial intelligence offer a solution to productivity stagnation in the United Kingdom? If so, how?

There are two ways in which worker productivity can rise. The first is an increase in the amount of capital per worker, which will raise output per hour worked, though at a diminishing rate for every additional unit of capital. The second, and more important for the long run, is innovation of both a technical and an operational nature, which raises the potential output from a given mix of capital and labour.

Artificial intelligence (AI) is a technical innovation which will itself spawn future technical and process innovations. In that sense, it is a general purpose technology like electricity and the internal combustion engine. Brynjolfsson et al. (2017) recently estimated that the application of AI to US transport and call centres alone would increase US average worker productivity by 0.27% per year over 10 years.

Another example is the use of machine learning technology in medical diagnosis. AI applications in this sphere will free up time for medical professionals to devote to other parts of patient treatment. They are also expected to increase the speed and accuracy of diagnosis (Hsu 2017).

How could the impact of AI on productivity best be measured?

The measurement of the economic impact of AI raises similar issues to the ones identified by Bean (2016) in his review of the UK's official statistics. Broadly, they fall into two categories:

- zero-priced services, e.g. Google Search, YouTube and Facebook, which have at least in part substituted for pre-digital products for which one had to pay;
- quality improvements which may be unaccounted for in statistics which only consider price changes over time.

AI may increase the share of online services which are zero-priced, by increasing the ability of platforms to better target advertisements and commercial products at users, thereby generating more revenue with which free services can be financed. More significantly, AI will increase the quality of a range of services, such as NHS treatment, in a way that will likely not be fully captured by ONS surveys.

AI will also facilitate the automation of a number of economic activities, notably driving, leading to more efficient resource utilisation and a commensurate decline in capital investment on vehicles (cf. Lilico and Sinclair 2016). It is likely that GDP statistics will capture the decline in car and truck production. The related improvement in resource utilisation will be reflected in total factor productivity statistics,¹⁴⁵³ but quality improvements such as more comfortable travel and greater safety will be only imperfectly reflected.

The best way to measure the impact of AI on productivity will continue to be as output per hour worked. However, we should be conscious that this measure will only partly capture the welfare benefits from AI. Some improvements to official measurements may be obtained by following Bean's (2016) recommendation of periodic quality adjustments to CPI – which would normally lead to lower inflation measures – as well as updated economic surveys aimed at the disintermediated economy.

How robust are predictions about job losses and job creation as a result of artificial intelligence and automation?

The problem with predictions about the impact of AI on employment is that economists and AI practitioners have some idea of the sectors which are vulnerable to automation – especially over the long run – but they have only a dim notion of the new jobs which AI will give rise to.

The problem is that automation will not only increase the demand for associated activities such as computer scientists and engineers. By raising the real incomes of consumers through lower prices and more efficient resource utilisation, AI will

¹⁴⁵³ Total factor productivity is measured as the residual of output growth after accounting for increases in capital and labour inputs. It is therefore intended as a crude measure of innovation.

also increase general purchasing power. The positive employment consequences of the latter effect are an order of magnitude greater than those from the direct impact of AI on demand for related skills.

But consumer tastes and preferences are ever-changing and highly unpredictable: who could have ventured in 1960 that the UK market would be served by more than 23,000 personal trainers (IBIS World 2017)? Given increasing economic dynamism, today we find ourselves in a similar position with regard to the job market of the 2050s.

Frey and Osborne (2013) famously forecast that 47 per cent of total US employment is presently vulnerable to automation. They further document a negative relationship between skills and propensity to automation which would raise important distributional and public policy consequences. But looking at job losses without considering the job gains is like examining a project's viability by only considering expected outlays and not expected revenue.

Autor (2015) provides a more sanguine picture which is less precise but probably more accurate in its predictions than those of Frey and Osborne. He predicts that the recent jobs polarisation – between low-skill manual jobs and high-skill intellectual occupations – will not persist with automation. He also offers the advent of the ATM in banks as an example of the counterintuitive impact of technology: by releasing bank tellers from low-value routine tasks, it increased their productivity, but it also led to an increase in branch numbers and thus in overall bank teller employment (cf. Bessen 2015).

However, it cannot be said that existing academic and official estimates of the employment effects of automation are mutually consistent nor robust.

Should the Government consider how to mitigate the potential impact of artificial intelligence on jobs?

There are two medium-term consequences of AI which government policy might be expected to address: the replacement by machines of tasks currently done by humans, and the change in the mix of skills required for remunerative employment in the economy.

The government already operates a number of schemes aimed at income support of the unemployed and underemployed. There are also retraining and reskilling programmes for those affected by deindustrialisation. To this must be added general education policy at the primary and secondary levels, as well as apprenticeships and higher education, all of which are partly or wholly subsidised by the taxpayer.

Before the government embarks on new programmes aimed at mitigating the impact of AI, it should establish what the impact will be and whether any mitigation is needed. But for the reasons outlined above, it is too early to estimate with any accuracy what the net effect on job creation will be. Additional intervention by government would be premature.

Should Government consider a pilot scheme for Universal Basic Income, or Universal Basic Services, as some other countries are currently trialling?

The universal basic income is a proposal to replace the raft of welfare schemes currently in place. Its advantages are budget transparency and the removal of administrative bureaucracy, as the required due diligence and surveillance could presumably be undertaken by a small number of people within HMRC.

However, UBI is financially viable only if it acts as a substitute for all existing programmes. Even then it would likely require substantial deregulation of planning rules to achieve the increases in productivity growth which would ensure its long-term sustainability. In the present context, UBI would be unaffordable so it is unrealistic to consider trialling it at this stage.

Universal basic services are so far a vague notion which would extend the principle of 'free at the point of use' from healthcare services to housing, food, transport and other key items of household expenditure (UCL 2017). There would be budgetary problems associated with assuring free provision of all these goods and services to everyone regardless of willingness and ability to pay. But surely the most important objection is that the record of poor performance and rationing by the NHS (cf. Niemietz 2016) cannot compare to the welfare gains achieved through competitive provision of housing, food and transport.¹⁴⁵⁴

What role should the Government take?

The role of government in AI should focus on easing the transition to new technologies and the processes enabled by machine learning and associated innovations. This involves:

- Facilitating the adoption of AI technologies in public and publicly regulated services, notably healthcare and public transport, where trade unions are likely to oppose mechanisation and demand that surplus workers be retained at great cost to the taxpayer.
- Removing labour market restrictions which lower the opportunity cost of AI adoption – by raising the cost of employment – and make it more difficult for workers to redeploy to sectors and activities which are less vulnerable to substitution by AI.
- Reducing capital taxation which discourages investment in new technologies, thereby slowing down innovation and curbing the UK's productivity growth potential.

¹⁴⁵⁴ A rigorous examination of problems in the British housing and transport market reveals that they are rooted in excessive regulation and misguided state intervention rather than any alleged failure of market processes (cf. Niemietz 2016; Wellings 2016).

- Encouraging the decentralised and competitive provision of education and skills. It is difficult to anticipate the demands of the labour market many decades into the future, but it is likely that the creative thinking of thousands of independent organisations will do a better job than centralised bureaucracy in Whitehall.

In addition, there should be efforts to improve and update data collection by official statistical bodies to ensure, firstly, that public authorities have accurate information about labour market and productivity trends, and secondly, that associated interventions such as monetary policy and public expenditure are responding to real needs in the economy, as would hopefully be reflected in properly compiled inflation indices and output figures.

References

Autor, D. H. (2015) Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives* 29.3 (Summer): 3-30. <https://economics.mit.edu/files/11563>

Bean, C. (2016) Independent review of UK economic statistics. Cabinet Office. <https://www.gov.uk/government/publications/independent-review-of-uk-economic-statistics-final-report>

Bessen, J. (2015) How computer automation affects occupations: technology, jobs, and skills. Law & Economics Working Paper No. 15-49. Boston University School of Law. <http://www.bu.edu/law/files/2015/11/NewTech-2.pdf>

Brynjolfsson, E., D. Rock, and C. Syverson. (2017) Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics. In Agrawal et al. (eds.), *Economics of artificial intelligence*. National Bureau of Economic Research. <http://www.nber.org/chapters/c14007.pdf>

Frey, C. B., and M. A. Osborne. (2013) The future of employment: how susceptible are jobs to computerisation? https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

Hsu, J. (2017) Can a crowdsourced AI medical diagnosis app outperform your doctor? *Scientific American*, 11 August. <https://www.scientificamerican.com/article/can-a-crowdsourced-ai-medical-diagnosis-app-outperform-your-doctor/>

IBIS World. (2017) Personal trainers in the UK: market research report. <https://www.ibisworld.co.uk/industry-trends/market-research-reports/arts-entertainment-recreation/personal-trainers.html>

Lilico, A., and M. Sinclair. (2016) The cost of non-Europe in the sharing economy. Research paper by Europe Economics. Brussels: European Parliamentary Research Service.

[http://www.europarl.europa.eu/RegData/etudes/STUD/2016/558777/EPRS_STU\(2016\)558777_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2016/558777/EPRS_STU(2016)558777_EN.pdf)

Niemietz, K. (2016) Universal healthcare without the NHS. IEA Hobart Paperback 185. London: Institute of Economic Affairs.

<https://iea.org.uk/publications/universal-healthcare-without-the-nhs/>

----- (2016) The housing crisis: a briefing. London: Institute of Economic Affairs. <https://iea.org.uk/publications/research/the-housing-crisis-a-briefing>

UCL. (2017) Social prosperity for the future: a proposal for Universal Basic Services. Social Prosperity Network Report: an IGP Knowledge Network. London: University College London.

https://www.ucl.ac.uk/bartlett/igp/sites/bartlett/files/universal_basic_services_-_the_institute_for_global_prosperity_.pdf

Wellings, R. (2016) Without delay: getting Britain's railways moving. IEA Discussion Paper 69. London: Institute of Economic Affairs.

<https://iea.org.uk/publications/research/without-delay-getting-britains-railways-moving>

11 December 2017