



HOUSE OF LORDS

Select Committee on Artificial Intelligence

Report of Session 2017–19

AI in the UK: ready, willing and able?

Ordered to be printed 13 March 2018 and published 16 April 2018

Published by the Authority of the House of Lords

Select Committee on Artificial Intelligence

The Select Committee on Artificial Intelligence was appointed by the House of Lords on 29 June 2017 “to consider the economic, ethical and social implications of advances in artificial intelligence.”

Membership

The Members of the Select Committee on Artificial Intelligence were:

<u>Baroness Bakewell</u>	<u>The Lord Bishop of Oxford</u>
<u>Lord Clement-Jones</u> (Chairman)	<u>Lord Puttnam</u>
<u>Lord Giddens</u>	<u>Viscount Ridley</u>
<u>Baroness Grender</u>	<u>Baroness Rock</u>
<u>Lord Hollick</u>	<u>Lord St John of Bletso</u>
<u>Lord Holmes of Richmond</u>	<u>Lord Swinfen</u>
<u>Lord Levene of Portsoken</u>	

Declaration of interests

See Appendix 1.

A full list of Members’ interests can be found in the Register of Lords’ Interests:

<http://www.parliament.uk/mps-lords-and-offices/standards-and-interests/register-of-lords-interests>

Publications

All publications of the Committee are available at:

<http://www.parliament.uk/ai-committee>

Parliament Live

Live coverage of debates and public sessions of the Committee’s meetings are available at:

<http://www.parliamentlive.tv>

Further information

Further information about the House of Lords and its Committees, including guidance to witnesses, details of current inquiries and forthcoming meetings is available at:

<http://www.parliament.uk/business/lords>

Committee staff

The staff who worked on this inquiry Committee were Luke Hussey (Clerk), Dr Ben Taylor (Policy Analyst) and Hannah Murdoch (Committee Assistant).

Contact details

All correspondence should be addressed to the Select Committee on Artificial Intelligence, Committee Office, House of Lords, London SW1A 0PW. Telephone 020 7219 4384. Email

HLAIAdHoc@parliament.uk

Twitter

You can follow the Committee on Twitter: [@LordsAICom](https://twitter.com/LordsAICom).

CONTENTS

	<i>Page</i>
Summary	5
Chapter 1: Introducing artificial intelligence	11
Our inquiry	12
Defining artificial intelligence	13
Figure 1: Definitions of artificial intelligence	13
Box 1: Common terms used in artificial intelligence	14
Figure 2: Deep neural networks	15
Categories of artificial intelligence	15
History	16
Recent reports	19
<i>Robotics and artificial intelligence</i>	19
<i>Machine learning: the power and promise of computers that learn by example</i>	20
<i>Data Management and Use: Governance in the 21st century</i>	20
<i>Growing the artificial intelligence industry in the UK</i>	20
<i>Industrial Strategy: Building a Britain fit for the future</i>	20
Impact on politics	21
Chapter 2: Engaging with artificial intelligence	22
General understanding, engagement and public narratives	22
Everyday engagement with AI	25
Chapter 3: Designing artificial intelligence	28
Access to, and control of, data	28
Anonymisation	31
Strengthening access and control	31
Box 2: Open Banking	34
Intelligible AI	36
Technical transparency	38
Explainability	39
Addressing prejudice	41
Data monopolies	44
Box 3: The Competition and Markets Authority	45
Chapter 4: Developing artificial intelligence	47
Investment in AI development	47
Box 4: Start-ups and SMEs	48
Figure 3: Investment rounds	50
Turning academic research into commercial potential	53
Box 5: What is a spin-out?	54
Improving access to skilled AI developers	55
Diversity of talent	57
Immigration and overseas skilled workers	58
Maintaining innovation	60
Chapter 5: Working with artificial intelligence	63
Productivity	63
Box 6: What is productivity?	63
Figure 4: UK job productivity	64
Box 7: Broadband speeds	65

Government adoption, and procurement, of artificial intelligence	66
Impact on the labour market	69
Figure 5: Percentage of working people employed in each industry group, 1901–2011	73
Figure 6: How worried are you that your job will be replaced by AI in the near future?	74
National Retraining Scheme	75
Chapter 6: Living with artificial intelligence	77
Education and artificial intelligence	77
Impact on social and political cohesion	81
Inequality	83
Chapter 7: Healthcare and artificial intelligence	87
The opportunity	87
The value of data	88
Box 8: DeepMind and the Royal Free London NHS Foundation Trust	90
Box 9: Caldicott Guardians	91
Using AI	92
Chapter 8: Mitigating the risks of artificial intelligence	95
Legal liability	95
Criminal misuse of artificial intelligence and data	98
Autonomous weapons	101
Box 10: UK Government definitions of automated and autonomous systems	102
Box 11: Definitions of lethal autonomous weapons systems used by other countries	104
Chapter 9: Shaping artificial intelligence	106
Leading at home	106
The AI Council and the Government Office for AI	106
Centre for Data Ethics and Innovation	107
Box 12: The Human Fertilisation and Embryology Authority	108
A National Institute for AI	109
Regulation and regulators	112
Assessing policy outcomes	116
A vision for Britain in an AI world	117
An AI Code	120
Box 13: Data Protection Impact Assessments	123
Summary of conclusions and recommendations	126
Appendix 1: List of Members and declarations of interest	139
Appendix 2: List of witnesses	141
Appendix 3: Call for evidence	153
Appendix 4: Historic Government policy on artificial intelligence in the United Kingdom	155
Appendix 5: Note of Committee visit to DeepMind: Wednesday 13 September 2017	167
Appendix 6: Note of Committee visit to Cambridge: Thursday 16 November 2017	169

Appendix 7: Note of Committee visit to BBC Blue Room: Monday 20 November 2017	176
Appendix 8: Note of SME Roundtable Event at techUK: Thursday 7 December 2017	177
Appendix 9: Recommendations relevant to the Government's new AI organisations	179
Appendix 10: Acronyms and glossary	180

Evidence is published online at <http://www.parliament.uk/ai-committee> and available for inspection at the Parliamentary Archives (020 7129 3074).

Q in footnotes refers to a question in oral evidence.

SUMMARY

Our inquiry has concluded that the UK is in a strong position to be among the world leaders in the development of artificial intelligence during the twenty-first century. Britain contains leading AI companies, a dynamic academic research culture, a vigorous start-up ecosystem and a constellation of legal, ethical, financial and linguistic strengths located in close proximity to each other. Artificial intelligence, handled carefully, could be a great opportunity for the British economy. In addition, AI presents a significant opportunity to solve complex problems and potentially improve productivity, which the UK is right to embrace. Our recommendations are designed to support the Government and the UK in realising the potential of AI for our society and our economy, and to protect society from potential threats and risks.

Artificial intelligence has been developing for years, but it is entering a crucial stage in its development and adoption. The last decade has seen a confluence of factors—in particular, improved techniques such as deep learning, and the growth in available data and computer processing power—enable this technology to be deployed far more extensively. This brings with it a host of opportunities, but also risks and challenges, and how the UK chooses to respond to these, will have widespread implications for many years to come. The Government has already made welcome advances in tackling these challenges, and our conclusions and recommendations are aimed at strengthening and extending this work.

AI is a tool which is already deeply embedded in our lives. The prejudices of the past must not be unwittingly built into automated systems, and such systems must be carefully designed from the beginning. Access to large quantities of data is one of the factors fuelling the current AI boom. We have heard considerable evidence that the ways in which data is gathered and accessed needs to change, so that innovative companies, big and small, as well as academia, have fair and reasonable access to data, while citizens and consumers can protect their privacy and personal agency in this rapidly evolving world.

To do this means not only using established concepts, such as open data and data protection legislation, but also the development of new frameworks and mechanisms, such as data portability and data trusts. Large companies which have control over vast quantities of data must be prevented from becoming overly powerful within this landscape. We call on the Government, with the Competition and Markets Authority, to review proactively the use and potential monopolisation of data by big technology companies operating in the UK.

Companies and organisations need to improve the intelligibility of their AI systems. Without this, regulators may need to step in and prohibit the use of opaque technology in significant and sensitive areas of life and society. To ensure that our use of AI does not inadvertently prejudice the treatment of particular groups in society, we call for the Government to incentivise the development of new approaches to the auditing of datasets used in AI, and to encourage greater diversity in the training and recruitment of AI specialists.

The UK currently enjoys a position as one of the best countries in the world in which to develop artificial intelligence, but this should not be taken for granted. We recommend the creation of a growth fund for UK SMEs working with AI to help them scale their businesses; a PhD matching scheme with the costs shared

between the private sector; and the standardisation of mechanisms for spinning out AI start-ups from the excellent research being done within UK universities. We also recognise the importance of overseas workers to the UK's AI success, and recommend an increase in visas for those with valuable skills in AI-related areas. We are also clear that the UK needs to look beyond the current data-intensive focus on deep learning, and ensure that investment is made in less researched areas of AI in order to maintain innovation.

Many of the hopes and the fears presently associated with AI are out of kilter with reality. While we have discussed the possibilities of a world without work, and the prospects of superintelligent machines which far surpass our own cognitive abilities, we believe the real opportunities and risks of AI are of a far more mundane, yet still pressing, nature. The public and policymakers alike have a responsibility to understand the capabilities and limitations of this technology as it becomes an increasing part of our daily lives. This will require an awareness of when and where this technology is being deployed. We recommend that industry, via the AI Council, establish a voluntary mechanism to inform consumers when artificial intelligence is being used to make significant or sensitive decisions.

AI will have significant implications for the ways in which society lives and works. AI may accelerate the digital disruption in the jobs market. Many jobs will be enhanced by AI, many will disappear and many new, as yet unknown jobs, will be created. A significant Government investment in skills and training is imperative if this disruption is to be navigated successfully and to the benefit of the working population and national productivity growth. This growth is not guaranteed: more work needs to be done to consider how AI can be used to raise productivity, and it should not be viewed as a general panacea for the UK's wider economic issues.

As AI decreases demand for some jobs but creates demand for others, retraining will become a lifelong necessity and pilot initiatives, like the Government's National Retraining Scheme, could become a vital part of our economy. This will need to be developed in partnership with industry, and lessons must be learned from the apprenticeships scheme. At earlier stages of education, children need to be adequately prepared for working with, and using, AI. For a proportion, this will mean a thorough education in AI-related subjects, requiring adequate resourcing of the computing curriculum and support for teachers. For all children, the basic knowledge and understanding necessary to navigate an AI-driven world will be essential. In particular, we recommend that the ethical design and use of technology becomes an integral part of the curriculum.

In order to encourage adoption across the UK, the public sector should use targeted procurement to provide a boost to AI development and deployment. In particular, given the impressive advances of AI in healthcare, and its potential, we considered the health sector as a case study. The NHS should look to capitalise on AI for the public good, and we outline steps to overcome the barriers and mitigate the risks around widespread use of this technology in medicine.

Within the optimism about the potential of AI to benefit the UK, we received evidence of some distinct areas of uncertainty. There is no consensus regarding the adequacy of existing legislation should AI systems malfunction, underperform or otherwise make erroneous decisions which cause harm. We ask the Law Commission to provide clarity. We also urge AI researchers and

developers to be alive to the potential ethical implications of their work and the risk of their work being used for malicious purposes. We recommend that the bodies providing grants and funding to AI researchers insist that applications for such funding demonstrate an awareness of the implications of their research and how it might be misused. We also recommend that the Cabinet Office's final Cyber Security & Technology Strategy consider the risks and opportunities of using AI in cybersecurity applications, and conduct further research as how to protect datasets from any attempts at data sabotage.

The UK must seek to actively shape AI's development and utilisation, or risk passively acquiescing to its many likely consequences. There is already a welcome and lively debate between the Government, industry and the research community about how best to achieve this. But for the time being, there is still a lack of clarity as to how AI can best be used to benefit individuals and society. We propose five principles that could become the basis for a shared ethical AI framework. While AI-specific regulation is not appropriate at this stage, such a framework provides clarity in the short term, and could underpin regulation, should it prove to be necessary, in the future. Existing regulators are best placed to regulate AI in their respective sectors. They must be provided with adequate resources and powers to do so.

By establishing these principles, the UK can lead by example in the international community. There is an opportunity for the UK to shape the development and use of AI worldwide, and we recommend that the Government work with Government-sponsored AI organisations in other leading AI countries to convene a global summit to establish international norms for the design, development, regulation and deployment of artificial intelligence.

“As soon as it works, no one calls it AI anymore ...”¹

You wake up, refreshed, as your phone alarm goes off at 7:06am, having analysed your previous night’s sleep to work out the best point to interrupt your sleep cycle.² You ask your voice assistant for an overview of the news, and it reads out a curated selection based on your interests.³ Your local MP is defending herself—a video has emerged which seems to show her privately attacking her party leader. The MP claims her face has been copied into the footage, and experts argue over the authenticity of the footage.⁴ As you leave, your daughter is practising for an upcoming exam with the help of an AI education app on her smartphone, which provides her with personalised content based on her strengths and weaknesses in previous lessons.⁵

On your way to work, your car dashboard displays the latest traffic information, and estimates the length of your journey to the office, based on current traffic conditions and data from previous journeys.⁶ On arrival, you check your emails, which have been automatically sifted into relevant categories for you.⁷ A colleague has sent you several dense legal documents, and software automatically highlights and summarises the points most relevant to a meeting you have later.⁸ You read another email, sent by your partner, asking if he can borrow your bank login details to quickly check something. On closer inspection you decide it is probably a fake, but still, you hesitate before deleting it, wondering briefly how the spammers captured his writing style so unerringly.⁹

-
- 1 Bertrand Meyer, ‘John McCarthy’, *Communications of the ACM* (28 October 2011): <https://cacm.acm.org/blogs/blog-cacm/138907-john-mccarthy/fulltext> [accessed 8 March 2018]
 - 2 A range of smartphone apps exist which can track sleep cycles by monitoring bed movements or snoring, and use machine learning to attempt to wake you up during lighter periods of sleep. See for example, Brenda Stolyar, ‘Sleep Cycle app for Android will soon allow users to track sleep using sound’, *Digital trends* (14 February 2018): <https://www.digitaltrends.com/mobile/sleep-cycle-app-android-update/> [accessed 8 March 2018]
 - 3 The Amazon Echo and Google Home devices are just two of the many home AI assistants currently on the market with this feature.
 - 4 Lyrbird.ai, a US-based start-up, has used an AI voice emulation system to replicate the voices of former US President Barack Obama and current President Donald Trump. Other AI software is allowing users to swap faces into pre-existing video footage with relatively little technical skill necessary.
 - 5 Software known as ‘Intelligent Tutor Systems’, such as Tabor, Carnegie Learning and Front Row, is increasingly being used to track a learner’s progress and provide them with lessons and personalised content based on this.
 - 6 Most digital map services in use today use machine learning to predict traffic flow speeds and provide an estimated time until arrival.
 - 7 Many email services in use today, including Google’s Inbox and Microsoft Outlook, use AI to categorise emails by type and priority.
 - 8 A range of ‘lawtech’ businesses have begun offering software which examines legal documents for relevant information, and can assist with the preparation of legal contracts.
 - 9 A recent report from 26 academic and industry experts warned of the malicious applications of AI, including mass ‘spear phishing’ attacks. Current spear phishing attacks, whereby fraudulent emails are personalised to an individual target, usually in a bid to steal sensitive information, currently require significant human labour, but by automating this process, these attacks could be scaled-up in the near future. See Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation and OpenAI, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (February 2018): https://img1.wsimg.com/blobby/go/3d82daa4-97fe-4096-9c6b-376b92c619de/downloads/1c6q2kc4v_50335.pdf [accessed 1 March 2018]

You have other things to worry about though, as you head to a hospital appointment. However, after a chest x-ray, you are surprised when the doctor sits you down immediately afterwards, explaining that you look to have a mild lung infection—you had expected it to take weeks before the results came back.¹⁰

Your relief is short lived—a notification on your phone warns you of suspicious activity detected on your bank account, which has been automatically stopped as a result.¹¹ You call the bank, and someone called Sarah picks up, and helps you order a replacement card. Except, you soon realise, Sarah is not human at all, just a piece of software which sounds just like a real person.¹² You are a little unnerved you did not realise more quickly, but still, it got the job done, so you do not particularly mind.

After a quick detour to the local supermarket, where the products on the shelves have all been selected automatically based on previous customer demand, current shopping trends and the likely weather that day, you drive home.¹³ On your way back, your car detects signs that you are feeling slightly agitated, and chooses some music you have previously found relaxing.¹⁴ After dinner, you and your partner watch a film suggested by your TV, which somehow strikes just the right note for both of your normally divergent tastes.¹⁵ After dozing off, your house, predicting you are asleep by now, turns off the bathroom light and turns on the washing machine, ready for another day.¹⁶

10 AI-powered radiology software is beginning to enter limited usage, which can automate aspects of x-ray analysis, and significantly reduce the amount of time needed to get useful results from scans.

11 Companies like MasterCard and Visa have been using machine learning algorithms to detect fraudulent patterns of spending in debit and credit cards, and automatically freeze cards in response.

12 A number of companies are using AI-powered chatbots, which can handle routine interactions with customers, and recently NatWest began experimenting with ‘Cora’, an in-branch AI personality, which can help with basic customer queries.

13 A number of UK supermarket chains are now using machine learning algorithms to better predict customer demand for particular products, cutting down on unnecessary waste and missed sales.

14 Emotion recognition is currently a significant area of growth in AI development, and a number of facial recognition companies have claimed that their systems have achieved human or near-human levels of emotion recognition.

15 Online film and TV streaming services, such as Netflix and BBC iPlayer have used machine learning algorithms to suggest what to watch based on previous viewing preferences and a range of other factors.

16 Smart home hubs, which can control a range of different smart systems such as lighting and home appliances, are becoming increasingly commonplace, and are using machine learning to detect and automate household functions based on personal habits and behaviour.

AI in the UK: ready, willing and able?

CHAPTER 1: INTRODUCING ARTIFICIAL INTELLIGENCE

*“We propose that a two-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer”.*¹⁷

1. So went the proposal for the first academic workshop on the newly minted field of ‘artificial intelligence’ in September 1955. The challenge of replicating human intelligence in a machine may not have been solved over the course of one New Hampshire summer, but 63 years later, aspects of that dream are beginning to fulfil their promise.
2. Over the years, this quest has produced its fair share of successes and failures, but the past decade has produced another wave of excitement, mostly centred on the use of artificial neural networks and deep learning algorithms. These techniques have allowed machines to consistently replicate human abilities, such as the visual recognition of objects and faces, which had hitherto proved resistant to conventional computing. The prospect of a wide variety of human abilities soon being replicable by machines has in turn generated both excitement and anxiety in equal measures, as predictions about the impact of artificial intelligence (AI) have rapidly multiplied.
3. Every sentence in the example at the start of this report described a world driven and mediated by AI, featuring applications in use today or which will be available imminently. AI has moved out of the realms of science fiction and into our everyday lives, working away unnoticed behind the scenes. Yet just as electricity, or steam power before it, started out with particular, often somewhat niche, uses prior to gradually becoming fundamental to almost all aspects of economic and social activity, AI may well grow to become a pervasive technology which underpins our daily existence. Electrification had many consequences: unprecedented opportunities for economic development, new risks of injury and death by electrocution, and debates over models of control, ownership and access, to name just a few. We can expect a similar process as AI technology continues to spread through our societies. This report will examine some of the implications, both good and bad, of what could prove to be a defining technological shift of our age.

¹⁷ John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (31 August 1955), p 1: <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf> [accessed 5 February 2018]

Our inquiry

4. The House of Lords appointed this Committee “to consider the economic, ethical and social implications of advances in artificial intelligence” on 29 June 2017.¹⁸ From the outset of this inquiry, we have asked ourselves, and our witnesses, five key questions:
 - How does AI affect people in their everyday lives, and how is this likely to change?
 - What are the potential opportunities presented by artificial intelligence for the United Kingdom? How can these be realised?
 - What are the possible risks and implications of artificial intelligence? How can these be avoided?
 - How should the public be engaged with in a responsible manner about AI?
 - What are the ethical issues presented by the development and use of artificial intelligence?

It is the answers to these questions, and others, on which we have based this report, our conclusions and our recommendations.
5. We issued our call for evidence on 19 July 2017, and received 223 pieces of written evidence in response. We took oral evidence from 57 witnesses during 22 sessions held between October and December 2017. We are grateful to all who contributed their time and expertise. The witnesses are shown in Appendix 2. The call for evidence is shown in Appendix 3. The evidence received is published online.
6. Alongside our oral evidence programme, we conducted a number of visits. On 13 September, we visited DeepMind, an artificial intelligence company based in King’s Cross, London. On 16 November, we visited Cambridge and met with the staff of Microsoft Research Lab Cambridge, and with two start-ups working with AI, Prowler.io and Healx. We also met academics at the Leverhulme Centre for the Future of Intelligence, an interdisciplinary community of researchers studying the opportunities and risks of AI over coming decades. On 20 November, we visited the BBC to meet staff of the Blue Room, a media technology demonstration team exploring ways audiences find, consume, create and interact with content. Finally, on 7 December, the Committee visited techUK to participate in a roundtable discussion with companies working with artificial intelligence in the United Kingdom. We are grateful to all concerned. Notes of all of these visits are contained in the appendices of this report.
7. In the course of our inquiry, we were trained by ASI Data Science on how to programme a neural network, and we received a private briefing from the National Cyber Security Centre. Our Chairman met, informally, with Carolyn Nguyen, Director of Technology Policy for Microsoft.

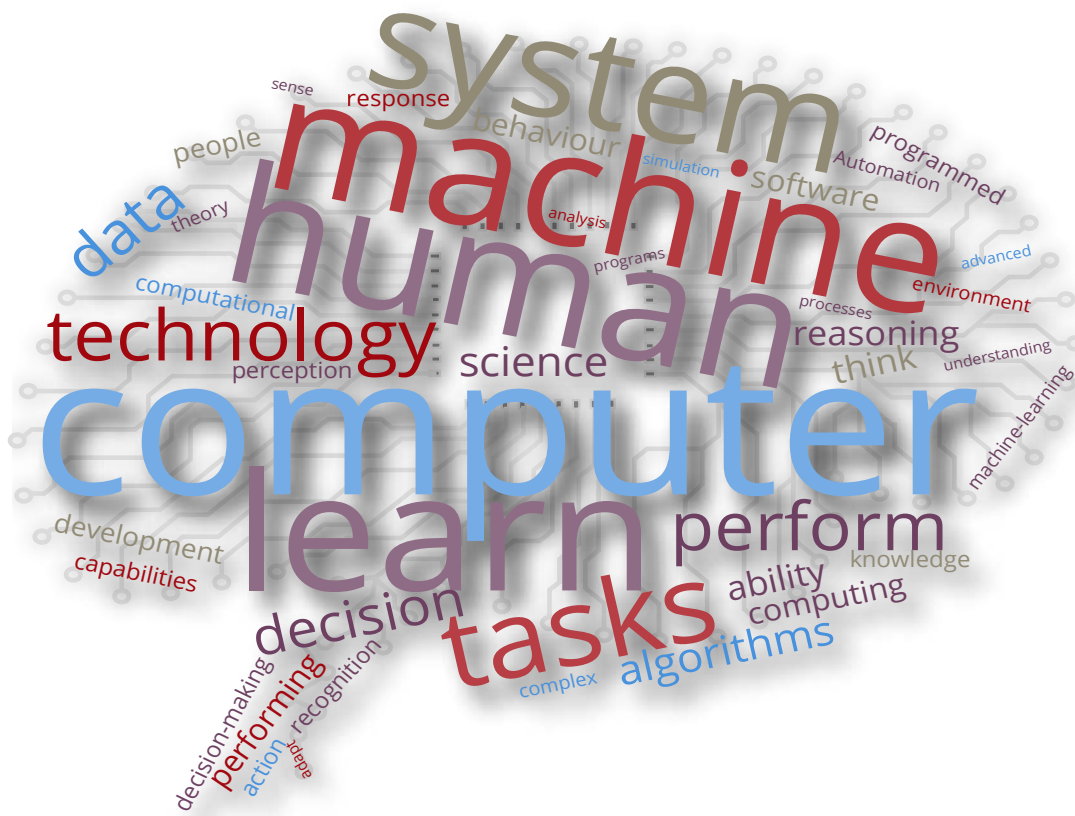
18 HL Deb, 29 June 2017, [col 562](#)

8. The members of the committee who carried out this inquiry are listed in Appendix 1, which shows our declared interests. Throughout the course of our inquiry we have been fortunate to have had the assistance of Dr Mateja Jamnik as our specialist adviser. We also engaged Angélica Agredo Montealegre, a PhD student at King’s College London, as another specialist adviser for part of the inquiry. Angélica was commissioned to research historic Government policy on artificial intelligence in the UK. This work has informed our recommendations, and we have published it as an appendix to this report. We are most grateful to them for their contribution to our work.

Defining artificial intelligence

9. There is no widely accepted definition of artificial intelligence.¹⁹ Respondents and witnesses provided dozens of different definitions. The word cloud (Figure 1) illustrates the diversity of the definitions we received, and shows the prominence of a few key words:

Figure 1: Definitions of artificial intelligence



¹⁹ This is perhaps unsurprising given the absence of any widely-accepted definition of organic intelligence, according to which AI is normally compared.

10. The debate around exactly what is, and is not, artificial intelligence, would merit a study of its own. For practical purposes we have adopted the definition used by the Government in its Industrial Strategy White Paper, which defined AI as:

“Technologies with the ability to perform tasks that would otherwise require human intelligence, such as visual perception, speech recognition, and language translation”.²⁰

11. Our one addition to this definition is that AI systems today usually have the capacity to learn or adapt to new experiences or stimuli. With this caveat, when we discuss AI in the following report, it is with this definition in mind.
12. Many technical terms are used in this field, the most common of which are summarised in Box 1.

Box 1: Common terms used in artificial intelligence

Algorithm

A series of instructions for performing a calculation or solving a problem, especially with a computer. They form the basis for everything a computer can do, and are therefore a fundamental aspect of all AI systems.

Expert system

A computer system that mimics the decision-making ability of a human expert by following pre-programmed rules, such as ‘if this occurs, then do that’. These systems fuelled much of the earlier excitement surrounding AI in the 1980s, but have since become less fashionable, particularly with the rise of neural networks.

Machine learning

One particular form of AI, which gives computers the ability to learn from and improve with experience, without being explicitly programmed. When provided with sufficient data, a machine learning algorithm can learn to make predictions or solve problems, such as identifying objects in pictures or winning at particular games, for example.

Neural network

Also known as an artificial neural network, this is a type of machine learning loosely inspired by the structure of the human brain. A neural network is composed of simple processing nodes, or ‘artificial neurons’, which are connected to one another in layers. Each node will receive data from several nodes ‘above’ it, and give data to several nodes ‘below’ it. Nodes attach a ‘weight’ to the data they receive, and attribute a value to that data. If the data does not pass a certain threshold, it is not passed on to another node. The weights and thresholds of the nodes are adjusted when the algorithm is trained until similar data input results in consistent outputs.

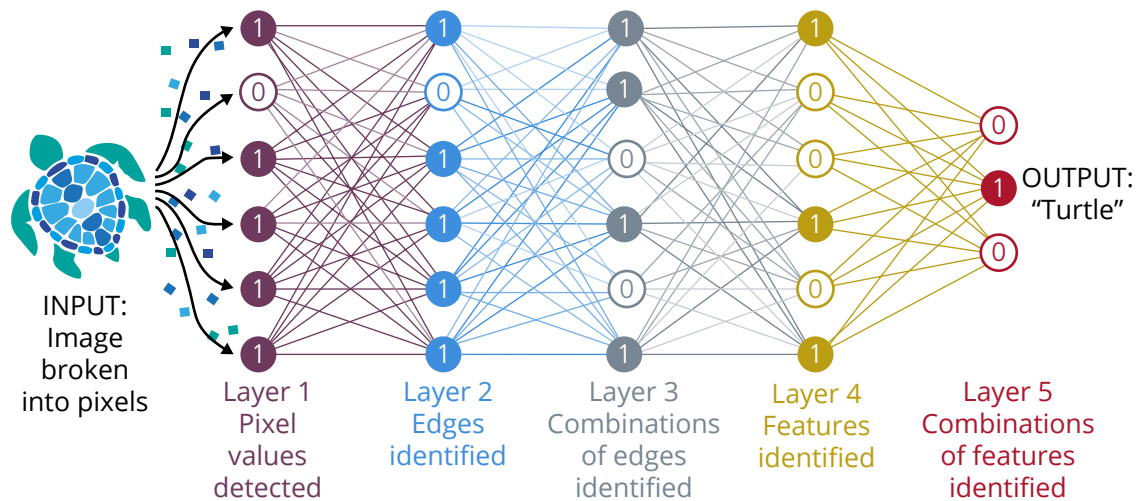
Deep learning

A more recent variation of neural networks, which uses many layers of artificial neurons to solve more difficult problems. Its popularity as a technique increased significantly from the mid-2000s onwards, as it is behind much of the wider interest in AI today. It is often used to classify information from images, text or sound (see Figure 2).

20 Department for Business, Energy and Industrial Strategy, *Industrial Strategy: Building a Britain fit for the future* (November 2017), p 37: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf [accessed 20 March 2018]

Figure 2: Deep neural networks

When data is fed into a deep neural network, each artificial neuron (labelled as “1” or “0” below) transmits a signal to linked neurons in the next level, which in turn are likely to fire if multiple signals are received. In the case of image recognition, each layer usually learns to focus on a particular aspect of the picture, and builds up understanding level by level.



Source: ‘New Theory cracks open the black box of deep neural networks’, *Wired* (10 August 2017): <https://www.wired.com/story/new-theory-deep-learning/> [accessed 8 March 2018]

13. We have also chosen to refer to the AI development sector, rather than an AI sector, on the grounds that it is mostly a particular sub-group within the technology sector which is currently designing, developing and marketing AI systems, but multiple sectors of the economy are currently deploying AI technology, and many more will likely join them in the near future.

Categories of artificial intelligence

14. Artificial intelligence can be viewed as ‘general’ or ‘narrow’ in scope. Artificial general intelligence refers to a machine with broad cognitive abilities, which is able to think, or at least simulate convincingly, all of the intellectual capacities of a human being, and potentially surpass them—it would essentially be intellectually indistinguishable from a human being.
15. Narrow AI systems perform specific tasks which would require intelligence in a human being, and may even surpass human abilities in these areas. However, such systems are limited in the range of tasks they can perform.
16. In this report, when we refer to artificial intelligence we are referring to narrow AI systems unless explicitly stated otherwise. It is these systems which have seen so much progress in recent years, and which are likely to have the greatest impact on our lives. By contrast, there has been little to no progress in the development of artificial general intelligence.²¹
17. The terms ‘machine learning’ and ‘artificial intelligence’ are also sometimes conflated or confused, but machine learning is in fact a particular type of artificial intelligence which is especially dominant within the field today. We are aware that many computer scientists today prefer to use ‘machine learning’ given its greater precision and lesser tendency to evoke misleading public perceptions.

21 Written evidence from Professor Michael Wooldridge ([AIC0174](#))

18. We have intentionally chosen to refer for the most part to artificial intelligence as a whole rather than machine learning. While this is partly because AI, for all its difficult baggage, is a far more familiar term to the public, it is mostly because we are aware that AI is a broad field. While machine learning is currently the most well-represented and successful branch, as the following historical overview illustrates, this has not always been the case, and may well not be the case in the future.
19. Many of the issues we will deal with are associated with the implications of machines which can simulate aspects of human intelligence. While in some cases the exact mechanisms by which they do this are of relevance, in many cases they are not, and we believe it important to retain a broad outlook on the societal impact of AI as a whole.

History

20. The field of artificial intelligence has been inextricably linked to the rise of digital computing, and many pioneers of the latter, such as Alan Turing and John McCarthy, were also closely involved with conceptualising and shaping the former. 1950 saw the publication of Turing's seminal paper, *Computing Machinery and Intelligence*, which helped to formalise the concept of intelligent machines and embed them within the rapidly growing field of digital computing.²²
21. Indeed, for better or worse, many of the concepts and terminology we still use to describe the field were bequeathed to us in this period. This included the concept of a 'Turing Test' to determine whether a machine has achieved 'true' artificial intelligence, and even the term 'artificial intelligence' itself. It was commonly thought at this time that the most promising way to achieve these ends was to mimic nature, which led to the first experiments with artificial 'neural networks' designed to very crudely approximate the networks of neurons in the human brain.
22. In the 1960s the field moved beyond the relatively small number of pioneers, mostly based in the United States and Britain, and the first major academic centres for AI research were established, at the Massachusetts Institute of Technology (MIT), Carnegie Mellon University, Stanford, and Edinburgh University. The period saw considerable enthusiasm for AI and its potential applications, with claims by some AI experts that the challenge of machine intelligence would be solved "within a generation".²³
23. By the 1970s these exuberant claims began to meet growing scepticism on both sides of the Atlantic. In the UK, discord within the AI research communities at Edinburgh and Sussex Universities prompted the Science Research Council to launch an inquiry, headed by Professor Sir James Lighthill, into the state of the field. The Lighthill Report of 1973, while supportive of AI research related to automation and computer simulations

22 A.M. Turing, 'Computing Machinery and Intelligence' in *Mind*, vol. 59, (1 October 1950), pp 433–460: https://www.cs.ox.ac.uk/activities/ieg/e-library/sources/t_article.pdf [accessed 5 February 2018]

23 This particular claim came from Marvin Minsky, an early pioneer of AI and noted sceptic of neural networks. However, Luke Muehlhauser of the Open Philanthropy Project has argued convincingly that some historical accounts of this period have over-exaggerated or misinterpreted the hyperbole of this period, and many computer scientists made far more moderate predictions. Luke Muehlhauser, 'What should we learn from past AI forecasts?', *Open Philanthropy Project* (September 2016): <https://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/what-should-we-learn-past-ai-forecasts> [accessed 5 February 2018]

of psychological and neurological processes, was deeply critical of much basic research into AI. It was also doubtful that general-purpose AI would be achievable within the twentieth century, if at all.²⁴ It is not in fact clear, as some have claimed, whether this led to a scaling back of research funding for AI, but scepticism towards and within the field certainly grew during this period, which is now referred to by many technologists as the first ‘AI winter’.²⁵ While some AI researchers criticised the manner in which the inquiry was conducted, and especially the fact that Lighthill was not himself a specialist in AI, the United States followed a similar trajectory, independent of Lighthill’s scepticism.

24. Despite these setbacks, research into AI continued, and by the 1980s some of this was starting to produce commercially viable applications. Among the first of these were ‘expert systems’, which sought to record and programme into machines the rules and processes used by specialists in particular fields, and produce software which could automate some forms of expert decision making, such as measuring the correct dosages when prescribing antibiotics.²⁶ By one contemporary estimate, at the end of the decade over half of Fortune 500 companies were involved in either developing or maintaining expert systems.²⁷ As primarily rule-based systems, these were mostly quite different from the machine learning systems of today, with little to no capacity to ‘learn’ new functionality.
25. The 1980s also saw the UK Government renew its troubled relationship with AI, with the ambitious Alvey Programme launched in 1983. Envisaged as a response to major state-sponsored computing research and development (R&D) projects elsewhere in the world, in particular Japan’s Fifth Generation, it sought to bring together researchers from the Government, universities and industry to investigate a range of issues, including AI. Despite overall funding of £350 million (equivalent to over £1 billion today) over four years, with £200 million coming directly from the Government, it failed in its central objective of improving the competitiveness of UK information technology (IT) businesses. The director of the programme, Brian Oakley, later claimed that too much emphasis was placed on pure R&D, at the expense of the development required to build viable products.²⁸
26. Disillusionment with the Alvey Programme coincided with a second global ‘AI winter’ at the end of the 1980s. Enthusiasm for expert systems waned as their limitations—high costs, requirements for frequent and time-consuming updates, and a tendency to become less useful and accurate as more rules

24 Science Research Council, *Artificial Intelligence: A paper symposium* (1973): http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm [accessed 5 February 2018]

25 As Muehlhauser has argued, a few AI researchers even remembered to this period as a “boomtime” for AI. Luke Muehlhauser, ‘What should we learn from past AI forecasts?’, *Open Philanthropy Project* (September 2016): <https://www.openphilanthropy.org/focus/global-catastrophic-risks/potential-risks-advanced-artificial-intelligence/what-should-we-learn-past-ai-forecasts> [accessed 5 February 2018]

26 Luke Dormehl, *Thinking Machines: The Quest for Artificial Intelligence—and where it’s taking us next*, 1st edition (New York: TarcherPerigee, an imprint of Penguin Random House LLC, 2017), p 30

27 Beth Enslow, ‘The Payoff from Expert Systems’, *Across the Board* (January/February 1989), p 56: <https://stacks.stanford.edu/file/druid:sb599zp1950/sb599zp1950.pdf> [accessed 1 March 2018]

28 Angeli Mehta, ‘Ailing after Alvey: The Alvey programme was Britain’s big chance to compete in information technology - Brian Oakley, a former director of the Alvey, reflects on what went wrong’, *New Scientist* (7 July 1990): <https://www.newscientist.com/article/mg12717242-300-ailing-after-alvey-the-alvey-programme-was-britains-big-chance-to-compete-in-information-technology-brian-oakley-a-former-director-of-alvey-reflects-on-what-went-wrong/> [accessed 31 January 2018]

were added—became apparent.²⁹ The Defence Advanced Research Projects Agency (DARPA), one of the main sources of government funding in the USA, decreased AI funding by a third between 1987 and 1989, while investment from the private sector also decreased.³⁰ The development of the internet and the World Wide Web also began diverting attention from AI R&D, offering as they did alternative models for organising, processing and disseminating information to that of AI systems.³¹

27. Even as the excitement and level of investment into expert systems and AI R&D more generally diminished, the late 1980s and 1990s saw AI applied to increasingly diverse functions, including predicting changes in the stock markets, data mining large corporate databases and developing visual processing systems like automatic number plate recognition cameras.³² Many of these new applications made less use of the rules and logic-based ‘symbolic AI’ approaches of previous decades. Instead, they deployed alternative machine learning approaches, which looked for statistical patterns and correlations in increasingly large datasets, paving the way for more recent developments in narrow AI systems.
28. Government support in the UK also continued, albeit in a drastically downscaled format, with the Department for Trade and Industry’s Neural Computing Technology Transfer Programme, which began in 1993 with a budget of £5.75 million, spread over six years. Intended to raise awareness of neural networks (now rebranded as ‘neural computing’) in business, the project encompassed an awareness campaign, and a demonstrator programme, which established seven clubs, managed and delivered by contracted consortia.³³ The subsequent evaluation claimed that “after 18 months 3,500 companies who had participated in awareness events could name an application area within their company where neural networks could be applied and 1,000 had taken some action to introduce applications”.³⁴ However, while the programme was thought to have provided “an important benefit in allowing companies to test neural networks in a low cost manner”, it failed to produce a legacy of subsidy-free investment in the technology.
29. The current wave of interest in AI was largely driven by developments in neural networks in the mid-2000s, when a team led by Geoffrey Hinton, a British researcher based at the University of Toronto, began to demonstrate the power of ‘deep learning’ neural networks. The team showed that these networks, which could automatically process unlabelled data, could be more effective at a wide range of tasks, such as image and speech recognition, than the more conventional algorithms then in use.

29 In short, the large quantities of tacit knowledge which many professions and experts relied on to do their jobs could overwhelm these rule-based systems. *Thinking Machines: The Quest for Artificial Intelligence—and where it’s taking us next*, p 32

30 *Thinking Machines: The Quest for Artificial Intelligence—and where it’s taking us next*, p 33. This waning enthusiasm was not universal though – the European ESPRIT project, and Japan’s Fifth Generation project, continued well into the 1990s.

31 Richard Susskind and Daniel Susskind, *The Future of the Professions: How technology will transform the work of human experts* (Oxford: Oxford University Press, 2015)

32 Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World* (New York: Basic Books, 2015), p 21

33 The National Archives, ‘Neural Computing Programme’: <http://webarchive.nationalarchives.gov.uk/20040117043522/http://www2.dti.gov.uk/iesea/aurep38b.html> [accessed 31 January 2018]

34 *Ibid.*

30. In the years since then, developments in AI in general, and deep learning in particular, have progressed rapidly. This is largely due to three factors: the growing availability of data with which to train AI systems, the continued growth in available computer processing power, and the development of more sophisticated algorithms and techniques. The widespread availability of cloud computing platforms, from Alibaba, Amazon and Microsoft in particular, has also helped by allowing clients to tap remotely into huge stores of computing power with relative ease, and without the need to maintain their own hardware. Finally, the growth of open source development platforms for AI—in particular Google’s TensorFlow, a library of components for machine learning—has reduced barriers to entry for researchers and commercial entities alike.
31. With this historic perspective in mind, there are three themes which we have considered throughout this report. Firstly, developments in the field of AI have been strongly characterised by boom and bust cycles, in which excitement and progress has been followed by disappointment and disillusionment. Otherwise known as the ‘AI winters’, researchers were unable to deliver on the full scale of their promises in a short enough time. While we believe, given the extent to which AI is now being used in actual commercial products and services today, that interest and investment is likely to persist this time round, we are also aware that the present excitement around AI is unlikely to be sustained indefinitely, and a note of caution would therefore be wise. Secondly, while AI research has been dominated by national government and universities since the 1950s, more recently this mantle has been passed to the private sector. While this has seen levels of investment increase to unprecedented levels, it also raises questions about the power and influence of the large tech companies—increasingly referred to as ‘Big Tech’—which are less fettered by the requirements for democratic accountability. We will return to this subject later in the report.³⁵ Thirdly, although the United States and the UK were early pioneers in AI research, China has invested heavily in the field, and aims to eclipse the efforts of other nations in the coming decades.

Recent reports

32. Ours is not the only recent report focusing on the impact of artificial intelligence. The following reports are those we have sought to build upon, and have borne in mind throughout the course of our inquiry.

Robotics and artificial intelligence

33. The House of Commons Science and Technology Committee published this report on 12 October 2016. The report recommended a greater emphasis on developing digital skills for the workforce of the future, and concluded that although it was too soon to set down specific regulations for the use of artificial intelligence, a standing commission on artificial intelligence should be created to examine its implications and establish principles to govern its development and applications.³⁶

³⁵ See Chapter 3.

³⁶ Science and Technology Committee, *Robotics and artificial intelligence* (Fifth Report, Session 2016–17, HC 145)

Machine learning: the power and promise of computers that learn by example

34. This report, published on 25 April 2017 by the Royal Society, focused on machine learning and was overseen by a working group drawn from academia and industry. The report outlined the major opportunities and challenges associated with the current trends in machine learning.³⁷ The report made a number of specific recommendations, suggesting that the Government should do more to promote open data standards, improve education and training in machine learning methods at all education levels, ensure that immigration and industrial strategy policies align with the needs of the UK AI development sector, and facilitate public dialogues on the opportunities and challenges of machine learning. Overall, the report argued in favour of specific sectoral approaches to regulating AI, rather than a more overarching, cross-sector approach.

Data Management and Use: Governance in the 21st century

35. The British Academy and Royal Society published this joint report on 28 June 2017. The report focused on the management of data, and concluded that while existing frameworks provided much of what is sufficient for today there is a need to develop a new framework to cope with future challenges. The report covered all aspects of data management and its use. This has been relevant to our work because of the importance of the management and use of data which feeds, and is generated by, artificial intelligence. The report recommended the establishment of a new body to steward the data governance landscape as a whole.

Growing the artificial intelligence industry in the UK

36. Professor Dame Wendy Hall, Regius Professor of Computer Science at the University of Southampton, and Dr Jérôme Pesenti, then CEO of BenevolentAI, chaired this review, staffed by civil servants, as part of the Government's Digital Strategy. The review, announced in March 2017, published its report on 15 October 2017. The Hall-Pesenti Review made 18 recommendations on how to make the UK the best place in the world for businesses developing AI. These recommendations focused on skills, increasing adoption of AI, ensuring data is used properly and securely, and building the UK's AI research capacity.

Industrial Strategy: Building a Britain fit for the future

37. This White Paper, published on 27 November 2017, set out the Government's long-term plan to boost productivity in the UK and strengthen the economy. The Strategy outlined four "Grand Challenges" for the UK, one of which was to put the country at the forefront of the artificial intelligence and data revolution. The Strategy also served as a response to the Hall-Pesenti Review. As such, the Strategy detailed a number of policies related to AI, and announced the establishment of a range of new institutions, which we discuss later in this report.

37 Royal Society, *Machine learning: the power and promise of computers that learn by example* (April 2017): <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf> [accessed 5 February 2018]

Impact on politics

38. Artificial intelligence will change the way we all relate to the world around us. The questions AI raises challenge existing ideological questions which have defined politics in the UK. The economy and society are changing, and all parties must stand ready to embrace and direct that change. As a cross-party committee, we recognise that in order for the UK to meet its potential and lead the way in shaping the future for society, AI policy must be committed to for the long-term, agreed by consensus and informed by views on all sides. It is in this spirit that we have made our report.

CHAPTER 2: ENGAGING WITH ARTIFICIAL INTELLIGENCE

39. The representation of artificial intelligence in popular culture is light-years away from the often more complex and mundane reality. Based on representations in popular culture and the media, the non-specialist would be forgiven for picturing AI as a humanoid robot (with or without murderous intentions), or at the very least a highly intelligent, disembodied voice able to assist seamlessly with a range of tasks. As discussed in the previous chapter, this is not a true reflection of its present capability, and grappling with the pervasive yet often opaque nature of artificial intelligence is becoming increasingly necessary for an informed society. This chapter focuses on the public's understanding of, and engagement with, AI and its implications, and how it can be improved.

General understanding, engagement and public narratives

40. Public perceptions of a subject as varied and amorphous as artificial intelligence will always be difficult to pinpoint with any precision, and are likely to change rapidly with every new innovation, scandal or accident which emerges into the public consciousness. AI is now such a wide-ranging subject that perceptions are increasingly dependent on who is using the technology, and to what purposes.³⁸ The Royal Society told us that their recent assessment of public attitudes towards machine learning in particular found that:

“ ... 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”.³⁹

41. Nevertheless, some of our witnesses suggested to us that while the British public is broadly aware of artificial intelligence, they often have inaccurate impressions of how it works, where it is to be found and its implications for them.⁴⁰ A survey specifically on machine learning, published on behalf of the Royal Society in April 2017, found that awareness of machine learning applications was relatively high, with 76% of respondents having heard of computers that can recognise speech and answer questions, and 89% being aware of at least one of the eight examples of machine learning used in the survey. However, it also found very limited awareness of how these applications worked, with just 9% of those surveyed having even heard the term ‘machine learning’, and 3% claiming that that they knew a great deal or fair amount about it.⁴¹

38 Written evidence from The Royal Society ([AIC0168](#))

39 *Ibid.*

40 Written evidence from Raymond Williams Foundation ([AIC0122](#)); Professor John Naughton ([AIC0144](#)); Baroness Harding of Winscombe ([AIC0072](#)); Google ([AIC0225](#)); Department of Computer Science University of Liverpool ([AIC0192](#)); Dr Toby Walsh ([AIC0078](#)) and Transport Systems Catapult ([AIC0158](#))

41 Ipsos MORI Social Research Institute, *Public views of Machine Learning: Findings from public research and engagement conducted on behalf of the Royal Society* (April 2017): <https://royalsociety.org/~media/policy/projects/machine-learning/publications/public-views-of-machine-learning-ipsos-mori.pdf> [accessed 5 February 2018]

42. Awareness and understanding of AI also varies across different segments of society. The Raymond Williams Foundation told us that “the more distant a person is from the subjects of science, technology, engineering and mathematics (especially statistics) ... the less likely s/he is to appreciate the changes that are underway”.⁴² Perhaps more surprisingly, Dr Ansgar Koene highlighted the fact that even young people were often surprised by the growing extent of AI in automated decision-making processes today.⁴³
43. Many AI researchers and witnesses connected with AI development told us that the public have an unduly negative view of AI and its implications, which in their view had largely been created by Hollywood depictions and sensationalist, inaccurate media reporting.⁴⁴ As well as unduly frightening people, witnesses said that these definitions were concentrating attention on threats which are still remote, such as the possibility of ‘superintelligent’ artificial general intelligence, while distracting attention away from more immediate risks and problems.⁴⁵
44. Such witnesses wanted a more positive take on AI and its benefits to be conveyed to the public, and feared that developments in AI might be threatened with the kind of public hostility directed towards genetically modified (GM) crops in the 1990s and 2000s.⁴⁶ In that case, companies and researchers struggled to articulate how they would benefit individual consumers, particularly in developed economies, and the perception took hold that large corporations would be the primary beneficiaries. Given that many of the efficiency and productivity benefits of AI adoption are likely to occur ‘behind the scenes’, and will not necessarily take the form of consumer-orientated products, there is a risk that GM-style opposition to AI could also grow. It is therefore up to businesses to ensure that benefits are passed on to consumers, in the form of innovative new AI-powered products, and cheaper and better services which are clearly linked to AI.
45. Journalists covering AI informed us that it was often difficult to cover AI and automation issues in a responsible and balanced manner, given the current level of public interest in the subject.⁴⁷ Sarah O’Connor, employment correspondent for the Financial Times, told us that “if you ever write an article that has robots or artificial intelligence in the headline, you are guaranteed that it will have twice as many people click on it”, and at least some journalists were sensationalising the subject in order to drive web traffic and advertising revenues.⁴⁸
46. However, there were those, both within and without AI development, who felt that in many cases AI developers and companies were at least partly responsible for public misunderstandings and confusion. Professor Kathleen Richardson and Nika Mahnič said that research scientists often inflated the potential of AI to attract prestigious research grants, resulting in “huge EU funded projects [which] are now promoting unfounded mythologies about

42 Written evidence from Raymond Williams Foundation ([AIC0122](#))

43 Written evidence from Dr Ansgar Koene ([AIC0208](#))

44 Written evidence from Dr Toby Walsh ([AIC0078](#)); Dr Will Slocombe ([AIC0056](#)); Ocado Group plc ([AIC0050](#)) and Foundation for Responsible Robotics ([AIC0188](#))

45 Written evidence from Professor John Naughton ([AIC0144](#)) and Dr Toby Walsh ([AIC0078](#))

46 Written evidence from Baroness Harding of Winscombe ([AIC0072](#)); University College London ([AIC0135](#)) and Royal Academy of Engineering ([AIC0140](#))

47 [Q 10](#) (Sarah O’Connor, Rory Cellan-Jones and Andrew Orłowski)

48 [Q 10](#) (Sarah O’Connor)

the capabilities of AI”.⁴⁹ Professor Sir David Spiegelhalter, President of the Royal Statistical Society, while noting the propensity for “utter puff” on AI from the media, maintained that ultimate responsibility for clarity lay with AI researchers and practitioners, and asked why they were not “working with the media and ensuring that the right sorts of stories appear”.⁵⁰ Other witnesses highlighted historical precedents, warning that excessive hyping had damaged the credibility of AI and AI researchers during earlier phases of its development, and there was a risk that these mistakes were being repeated in the present.⁵¹

47. There was also a wide variety of views as to what the intended purpose of public engagement in relation to AI should be. As mentioned, many AI researchers were concerned that the public were being presented with overly negative or outlandish depictions of AI, and that this could trigger a public backlash which could make their work more difficult. They told us that public engagement should be about building trust in AI to prevent this from happening.⁵²
48. Other witnesses warned against simplistically attempting to build trust in AI, as at least some applications of AI would not be worthy of trust. For example in cases where the technology may be used to mislead or deceive users, citizens and consumers would need the skills to decide whether to trust it for themselves.⁵³ Professor David Edgerton, Hans Rausing Professor of the History of Science and Technology, King’s College London, warned against Parliament or the Government attempting to “ensure that people embrace changes that are dictated from above”.⁵⁴ In his view their responsibility should be to “give people choices over which they can exercise their collective judgment. We should not assume that the stories that we tell about AI will reflect what will come to pass”.⁵⁵
49. One set of choices which members of the public will need to face is how personal data is used and, in some cases, abused. While we cannot know with certainty what shape AI will take in the future, it is highly likely that data will continue to be important. A number of witnesses believed that AI provided added impetus to the need to better educate the public on the use of their data and the implications for their privacy.⁵⁶ Professor Peter McOwan, Vice Principal (Public Engagement and Student Enterprise), Queen Mary University of London, told us that AI systems have become better at automatically combining separate datasets, and can piece together much more information about us than we might realise we have provided. He cited as an example ‘pleaserobme.com’, a short-lived demonstration website, which showed how publicly accessible social media data could be combined to automatically highlight home addresses of people who were on holiday.⁵⁷

49 Written evidence from Professor Kathleen Richardson and Ms Nika Mahnič ([AIC0200](#))

50 [Q 216](#) (Professor Sir David Spiegelhalter)

51 Written evidence from Dr Jerry Fishenden ([AIC0028](#))

52 Written evidence from Transport Systems Catapult ([AIC0158](#)); Dr Malcolm Fisk ([AIC0012](#)) and Innovate UK ([AIC0220](#))

53 Written evidence from Dr Ozlem Ulgen ([AIC0112](#)); [Q 123](#) (Dr Julian Huppert) and [Q 216](#) (Professor Sir David Spiegelhalter)

54 [Q 214](#) (Professor David Edgerton)

55 *Ibid.*

56 Written evidence from Big Brother Watch ([AIC0154](#)); Royal Academy of Engineering ([AIC0140](#)); Information Commissioner’s Office ([AIC0132](#)) and [Q 220](#) (Professor Sir David Spiegelhalter)

57 [Q 218](#) (Professor Peter McOwan)

50. **The media provides extensive and important coverage of artificial intelligence, which occasionally can be sensationalist. It is not for the Government or other public organisations to intervene directly in how AI is reported on, nor to attempt to promote an entirely positive view among the general public of its possible implications or impact. Instead, the Government must understand the need to build public trust and confidence in how to use artificial intelligence, as well as explain the risks.**

Everyday engagement with AI

51. Beyond a general awareness of AI in the abstract, the average citizen is, and will be increasingly, exposed to AI-enabled products and services in their day-to-day existence, often with little to no knowledge that this is the case. Although an improved general awareness and knowledge of AI and its implications may be desirable, some witnesses argued that there were limits to what the public could be reasonably expected to learn, or needed to know, with regards to an often highly technical subject. As such, they believed there should be a focus on particular aspects, scenarios and implications of AI and associated technology, rather than AI in the abstract. The Information Commissioner's Office (ICO) stated that there was a "need to be realistic about the public's ability to understand in detail how the technology works", and it would be better to focus on "the consequences of AI, rather than on the way it works", in a way that empowered individuals to exercise their rights.⁵⁸
52. Witnesses also made the point that, while consumers often had relatively few AI-specific concerns for now, they were gradually becoming more aware of the algorithmic nature of particular products and services, and the role of data in powering them. Colin Griffiths, of Citizens Advice, explained that while AI was enabling new products and services for consumers, particularly with regards to tailoring them to individual consumers, it was also enabling new things to be done to consumers which might not be in their interests, and which they might not be comfortable with.⁵⁹ Will Hayter, Project Director of the Competition and Markets Authority, agreed:
- “ ... the pessimistic scenario is that the technology makes things difficult to navigate and makes the market more opaque, and perhaps consumers lose trust and disengage from markets. The more optimistic scenario is that the technology is able to work for consumers”.⁶⁰
53. A number of witnesses highlighted the need for transparency when AI was being used with respect to consumers. For example, the Electronic Frontier Foundation told us of the need for transparency regarding the use of AI for dynamic or variable pricing systems, which allow businesses to vary their prices in real time.⁶¹ While this is mostly used at present to adjust prices in accordance with market fluctuations, as with online flight booking sites, it is increasingly allowing retailers to adjust prices according to what a specific individual customer is willing or able to pay, without necessarily making it apparent how much other customers are paying for the same thing.

58 Written evidence from Information Commissioner's Office ([AIC0132](#))

59 [Q 86](#) (Colin Griffiths)

60 [Q 86](#) (Will Hayter)

61 Written evidence from Electronic Frontier Foundation ([AIC0199](#))

54. While witnesses acknowledged the limits to consumer education, and noted that additional consumer protections may be necessary, many witnesses felt that mechanisms for informing consumers about the use of AI should be explored.⁶² The Market Research Society told us that “consumer facing marks with consumer recognition” could be a “useful tool in building consumer trust across markets and can form a vital part of the framework for regulating the use of AI”, a view shared by several other witnesses.⁶³ Professor Spiegelhalter said there might be certain difficulties in terms of defining what AI meant for these purposes, but was also broadly supportive.⁶⁴ However, Will Hayter was more sceptical of such an approach, pointing to the relative obscurity of the AdSense logo—intended to identify online adverts which have specifically been targeted at individuals—as a discouraging precedent.⁶⁵
55. Other witnesses were supportive of telling consumers when they were dealing with AI, especially with regards to chatbots, which have gradually begun to replace some forms of online customer service previously performed by humans. Future Intelligence highlighted how on some websites and services, AI chatbots inform the user: “I’m just a bot but I’ll try to find the answers for you”.⁶⁶ Doteveryone suggested that organisations using AI in this way should be required to declare where and how they use it, “similar to declarations of the use of CCTV or telephone recording”, and should be ready to explain AI functions and outcomes in plain terms.⁶⁷ With respect to the legal sector in particular, journalist and author Joanna Goodman, and academics from the University of Westminster Law School, told us that legal services using AI 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”.⁶⁸
56. The need for a more context-specific approach to informing people about AI is clear. While general improvements in public understanding of, and engagement with, AI are good, many people will still be unable or unwilling to grapple with this in the abstract. Nor, indeed, should they have to. They should be informed, in clear, non-technical language, in a way which allows them to make a reasoned decision regarding their own best interests.
57. Whatever the organisation or company in question, people should be provided with relevant information on how AI is involved in making significant or sensitive decisions about them, at the point of interaction. Without this, it will be very difficult for individuals and consumers to understand why they are presented with different information, offers and choices from their fellow citizens, to understand when they are being treated fairly or unfairly, and challenge decisions.

62 Written evidence from Dr Jerry Fishenden ([AIC0028](#)); Professor John Preston ([AIC0014](#)) and Dr Will Slocombe ([AIC0056](#))

63 Written evidence from Market Research Society ([AIC0130](#)); Deloitte ([AIC0075](#)) and Future Intelligence ([AIC0216](#))

64 [Q 217](#) (Professor Sir David Spiegelhalter)

65 [Q 87](#) (Will Hayter)

66 Written evidence from Future Intelligence ([AIC0216](#))

67 Written evidence from Doteveryone ([AIC0148](#))

68 Written evidence from Joanna Goodman, Dr Paresh Kathrani, Dr Steven Cranfield, Chrissie Lightfoot and Michael Butterworth ([AIC0104](#))

58. **Artificial intelligence is a growing part of many people's lives and businesses. It is important that members of the public are aware of how and when artificial intelligence is being used to make decisions about them, and what implications this will have for them personally. This clarity, and greater digital understanding, will help the public experience the advantages of AI, as well as to opt out of using such products should they have concerns.**
59. **Industry should take the lead in establishing voluntary mechanisms for informing the public when artificial intelligence is being used for significant or sensitive decisions in relation to consumers. This industry-led approach should learn lessons from the largely ineffective AdChoices scheme. The soon-to-be established AI Council, the proposed industry body for AI, should consider how best to develop and introduce these mechanisms.**

CHAPTER 3: DESIGNING ARTIFICIAL INTELLIGENCE

60. Large quantities of data have been an essential component of most contemporary advancements in AI. While there are AI algorithms which require smaller quantities of data, by and large the evidence we received anticipated no let-up in demand in the foreseeable future. The issues surrounding data, both in relation to digital technology and AI specifically, are complex and intertwined. In this chapter we consider issues to do with designing and developing artificial intelligence systems, including the use and monopolisation of data, the transparency and intelligibility of decisions made by AI systems, and the risk of prejudice in those decisions.
61. Two points must be explained in advance. Firstly, there is an important distinction to be made between data more generally, and personal data. While ‘data’ could refer to almost any information (such as temperature readings) on a computer, or which is intended to be held on a computer, ‘personal data’ has a specific meaning under the Data Protection Act 1998, and generally covers any set of information relating to individuals. While the balance of power and commercial opportunity afforded to companies and organisations will generally be determined by the quantity and quality of data they have access to, questions of privacy and personal agency relate more specifically to personal data.
62. Secondly, while we initially considered this area in terms of data ownership, after taking evidence from a number of experts in the field we came to believe that data control was a more appropriate concept. It was pointed out that custody and control of data were far more established legal concepts than ownership of data,⁶⁹ and Olivier Thereaux, Head of Technology, The Open Data Institute, also spelled out the conceptual difficulties inherent to data ownership:

“Data has a few qualities that makes it incompatible with notions of ownership. I can hold it, you can hold it, and my holding of it does not impact the value you derive from it ... Take the data about a phone call. I make a phone call to a friend. The data about that call has me as the data subject, but it could not easily be owned just by me because there are other data subjects. The person to whom I made that phone call is a data subject, the companies through which I made that phone call are another kind of data subject, a secondary data subject, and so on”.⁷⁰

We have accordingly decided to refer to data control, rather than data ownership, in this report.

Access to, and control of, data

63. Our witnesses painted a picture in which large technology corporations, headquartered in the United States but with a global reach, are accruing huge quantities of data, which is increasingly giving them a massive advantage in the development and application of AI systems when compared with smaller competitors and public sectors across the world.

69 [Q 56](#) (Dr Mercedes Bunz, Elizabeth Denham, Dr Sandra Wachter) and [Q 66](#) (Javier Ruiz Diaz, Olivier Thereaux, Frederike Kaltheuner). Javier Ruiz Diaz noted the precedent established in the *Your Response Limited vs Datateam Business Media Limited* [2014] EWCA Civ 281 (14 March 2014) case, when the Court of Appeal found that databases do not constitute tangible property of a kind which is capable of possession.

70 [Q 66](#) (Olivier Thereaux)

64. Professor Richard Susskind OBE told us of the “unprecedented concentration of wealth and power in a small number of corporations” such as Alibaba, Alphabet, Amazon, Apple, Facebook, Microsoft and Tencent, a view widely held among a variety of witnesses.⁷¹ Innovate UK noted that these “vast volumes of data” were increasingly allowing these companies to unlock the potential value in the AI systems they are developing.⁷² Alongside personal data, this also included many other disparate forms of data, which “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”, and it was the combination of various different kinds of data, often “largely unnoticed”, which was creating “significant commercial opportunities” for these corporations.⁷³ In some cases, as with DeepMind’s controversial deal with the Royal Free London NHS Foundation Trust (see Box 8), these companies are also making bespoke arrangements with public institutions to augment their already impressive access to data.
65. Several witnesses pointed to the network effects at work in information-based industries, which tend towards ‘winner-takes-all’ markets, and contributed to the growing dominance of these large technology companies.⁷⁴ As Dr Mike Lynch told us:
- “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 an advanced competitively defensive position”.⁷⁵
66. Some witnesses noted that the Government, and particular public institutions also had access to a wide range of datasets.⁷⁶ The NHS was frequently cited as holding some of the world’s largest and most comprehensive collections of health records, but there were also less commonly mentioned examples, such as Ordnance Survey, who told us of their “abundance of labelled data”, which included “rich data describing land parcels and geospatial features such as buildings, roads and railways”.⁷⁷ However, these often came with their own problems; many NHS records are still in paper form and require expensive and time-consuming digitisation before AI can make use of them, while Ordnance Survey observed that a lack of processing power is currently one of their major challenges in developing AI.⁷⁸
67. Meanwhile, many universities, charities, start-ups and small and medium sized enterprises (SMEs) complained that they could not easily access large, good quality datasets, and were struggling to compete with larger players as a result. Dr Mercedes Bunz, Senior Lecturer, Communications and

71 Written evidence from Professor Richard Susskind ([AIC0194](#)); Doteveryone ([AIC0148](#)); Professor Michael Wooldridge ([AIC0174](#)); Royal Academy of Engineering ([AIC0140](#)) and Transport Systems Catapult ([AIC0158](#))

72 Written evidence from Innovate UK ([AIC0220](#))

73 *Ibid.*

74 A conventional network effect is when a good or service becomes more useful as more people use it. For example, a telephone system becomes more useful as more people buy telephones. A data network effect produces similar benefits, but does so because machine learning allows a product or service to improve as more data is added to the network, which in turn draws more customers in, who in turn provide more data. Written evidence from The Economic Singularity Supper Club ([AIC0058](#)); Research Councils UK ([AIC0142](#)); Digital Catapult ([AIC0175](#)) and Dr Toby Walsh ([AIC0078](#))

75 Written evidence from Dr Mike Lynch ([AIC0005](#))

76 Braintree ([AIC0074](#)); Bikal ([AIC0052](#)); Charities Aid Foundation ([AIC0042](#)); Dr Mercedes Bunz ([AIC0048](#)); Doteveryone ([AIC0148](#)) and The Royal Society ([AIC0168](#))

77 Written evidence from Ordnance Survey ([AIC0090](#))

78 [Q 120](#) (Dr Julian Huppert) and written evidence from Ordnance Survey ([AIC0090](#))

Media Research Institute at the University of Westminster, emphasised how expensive the creation of datasets can be. She explained how ImageNet, one of the largest visual databases in the world, “employed at its peak 48,940 people in 167 countries”, who sorted around 14 million images.⁷⁹

68. The Charities Aid Foundation informed us that “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”.⁸⁰ Our evidence also suggests that many start-ups struggle to gain access to data. In some cases, this is because they need to develop a service before they can attract the customers who would in turn provide data, while in others it is because small start-ups cannot demonstrate their credibility to public institutions and organisations.⁸¹ With respect to small businesses, Kriti Sharma, Vice President of Artificial Intelligence and Bots, Sage, told us that “some 55% of small businesses are still using pen and paper, Excel spreadsheets, fragmented datasets”, which prevented them from making any substantive use of AI, although Sage also believed that developments in cloud technology were partly offsetting this by opening up new sources of data for SMEs.⁸² We were also told that, in many cases, even larger companies are unable or unwilling to make use of the data they have, with many different datasets scattered across different ‘siloes’ across the company.⁸³
69. All indications suggest that the present status quo will be disrupted to some extent by the upcoming General Data Protection Regulation (GDPR), which the UK is planning to adhere to regardless of the outcome of Brexit, and the new ePrivacy Regulation, which both come into force across the EU from 25 May 2018. With respect to data access, the GDPR’s introduction of a right to data portability is probably the most significant feature. While subject to some restrictions,⁸⁴ in many cases this will require companies and organisations to provide a user with their personal data in a structured, commonly used and machine readable form, free of charge. The intention is that consumers will be able to take their personal data from one service and, relatively seamlessly, transfer it to another, helping to prevent the ‘lock in’ effect which can dissuade customers from switching between service providers. A number of witnesses welcomed the new right to data portability. Dr Sandra Wachter, Postdoctoral Researcher in Data Ethics and Algorithms, Oxford Internet Institute, argued that it could “enhance competition in a very healthy way”, and facilitate access to data for new start-ups looking to compete with established giants.⁸⁵ However, Dr Bunz, while broadly welcoming the initiative, warned that on its own, “individual portability will not be sufficient to collect a dataset that allows the creation of knowledge and businesses”, and the UK was still in need of “a strategy to actively create big data, especially in areas of government interest such as healthcare, transport, science and education”.⁸⁶

79 Written evidence from Dr Mercedes Bunz ([AIC0048](#))

80 Written evidence from Charities Aid Foundation ([AIC0042](#))

81 Written evidence from UK Computing Research Committee ([AIC0030](#)); Michael Veale ([AIC0065](#)) and University College London (UCL) ([AIC0135](#))

82 [Q 77](#) (Kriti Sharma)

83 Written evidence from Bikal ([AIC0052](#))

84 Namely, the right to data portability only applies to personal data that an individual has provided to a controller, where the processing is based on the individual’s consent or for the performance of a contract, and when the processing is carried out by automated means.

85 [Q 59](#) (Dr Sandra Wachter)

86 Written evidence from Dr Mercedes Bunz ([AIC0048](#))

Anonymisation

70. Where the sharing of large quantities of personal data is concerned, another issue to be considered is how to make maximum use of this data, with the minimum possible infringement on the privacy of individuals. In practise, much of this comes down to an area of data science known as ‘anonymisation’ or ‘de-identification’, whereby datasets are processed in order to remove as much data which relates to individuals as possible, while retaining the usefulness of the dataset for the desired purpose. For example, a dataset of x-rays might be stripped of names, addresses and other identifying features, but which would still make the dataset useful for understanding trends in the development of particular diseases. This is now a routine process in the handling of many different kinds of datasets containing personal data.
71. However, some of our witnesses argued that de-identifying datasets is far less effective than often supposed. Olivier Thereaux of the Open Data Institute said “AI is particularly problematic, because it can be used extremely efficiently to re-identify people”, even in the face of “pretty good de-identification methods”.⁸⁷ De-identification often means removing identifying features, and AI systems can be effective at adding these back in, by, for example, cross-referencing other available datasets. It should also be noted that in at least some cases, there is a necessary trade-off—if more information is stripped out of a dataset in a bid to preserve privacy, this may also limit the potential usefulness and flexibility of the dataset in the process.⁸⁸
72. Elizabeth Denham, the Information Commissioner, told us this was an unrealistic view, stating: “I get frustrated when people say there is no such thing as perfect anonymisation; there is anonymisation that is sufficient”. She also noted provisions in the Data Protection Bill which will criminalise re-identification unless for particular legitimate purposes.⁸⁹ The Office of the National Data Guardian took a similar view that the public are broadly supportive of the use of anonymised data in the health sector if it was for clear health and social care purposes.⁹⁰ While we accept that de-identification methods have their flaws, and will always have to be kept up-to-date in the face of advances in AI, we are satisfied that for most purposes they offer an acceptable level of security and privacy. The ICO is closely monitoring this issue and has demonstrated a willingness to intervene as, and when, necessary. We welcome the ICO’s engagement in this area.

Strengthening access and control

73. Given the current data landscape, and the forthcoming new data protection regulations, it is perhaps unsurprising that we received a wide variety of opinions on how the situation might be improved.
74. Some of our witnesses, mostly larger technology companies, appeared to be broadly satisfied with the current mix of large repositories of privately-held data, bespoke agreements between particular companies and public organisations, and a patchwork of open data sources.⁹¹ A number of our

87 [Q 71](#) (Olivier Thereaux)

88 *Ibid.*

89 [Q 61](#) (Elizabeth Denham)

90 Written evidence from the National Data Guardian for Health and Care ([AIC0143](#))

91 Written evidence from CBI ([AIC0114](#)); The Economic Singularity Supper Club ([AIC0058](#)); The Law Society of England and Wales ([AIC0152](#)) and IBM ([AIC0160](#))

witnesses argued that this situation, while not directly recompensing individuals for their data (or in many cases indirectly), still delivered significant benefits to the wider population through the development of innovations, which might not be possible otherwise.⁹²

75. Those who held this position were also keen to emphasise, as the Confederation of British Industries (CBI) put it, that data “is not a finite resource”, and that unlike conventional assets, the use of a dataset for commercial purposes “does not inhibit the use of the same data for social or non-commercial purposes”.⁹³ They contested the very concept of ‘data monopolies’, arguing that “data has no intrinsic value in its raw sense”, and that in reality it is “the expertise in processing and application of data that creates value for organisations”. They were resistant to any idea of interfering with contractual freedoms, and told us that businesses who invest capital into building, maintaining, and protecting datasets have a right to seek a return on investment if they carry the reinvestment costs.⁹⁴ A number of other witnesses cautioned against undue interference, especially through regulation, and held that current data protection and competition law was more than adequate.⁹⁵
76. A number of witnesses advocated for more free and open access to public datasets, as part of the open data agenda, in order to help address the risks of data monopolies. This movement, supported by the Open Data Institute in the UK, believes that as much data as possible should be made freely available for use and reuse. UCL drew attention to data.gov.uk, which provides free access to data from Government departments and public bodies.⁹⁶ Several witnesses also applauded Transport for London’s efforts to make their data, which includes real-time transport information such as tube and bus departures, freely available, which has in turn led to the development of apps such as Citymapper.⁹⁷ Research Councils UK argued that, subject to legal, ethical and commercial restraints, all publicly funded research data is “a public good and should be made available with as few restrictions as possible”.⁹⁸ A number of witnesses felt that this was a clear area of data policy in which Government could, and should, intervene in order to encourage both the release of more public datasets and to further develop the open data standards, potentially encouraging others to do the same.⁹⁹
77. While many witnesses believed this would help to facilitate innovation in AI and develop a more level playing field, even open data was not without its critics. While several witnesses simply believed it was not sufficient on its own to counteract larger privately-held datasets, Dr Lynch argued that the open data movement was a “rather academic debate”, which does not account for the economic or strategic value of publicly-held datasets. He highlighted the perversity of unique forms of publicly-held data, particularly those held by the NHS, being given away for free to AI companies which may then hold those same organisations “to ransom” when selling these systems back to the public sector. In his view, the Government should take proper consideration

92 [Q 50](#) (David Kelnar and Eileen Burbidge)

93 Written evidence from CBI ([AIC0114](#))

94 *Ibid.*

95 Written evidence from The Economic Singularity Supper Club ([AIC0058](#)); The Law Society of England and Wales ([AIC0152](#)) and IBM ([AIC0160](#))

96 Written evidence from University College London ([AIC0135](#))

97 Written evidence from Accenture UK Limited ([AIC0191](#)) and University College London ([AIC0135](#))

98 Written evidence from Research Councils UK ([AIC0142](#))

99 Written evidence from Dr Jerry Fishenden ([AIC0028](#)); Braintree ([AIC0074](#)); Google ([AIC0225](#)); DeepMind ([AIC0234](#)) and Balderton Capital (UK) LLP ([AIC0232](#))

of the ‘strategic data’ it holds, and, for example, insist on “favoured nation pricing” in the contractual arrangements for supplying such data. Indeed, a number of witnesses were highly critical of the current tendency for bespoke public-private data deals, which they believed were allowing data to flow from the public sector to the private sector without securing proper value for the taxpayer.¹⁰⁰

78. Finally, there were those who believed that far more emphasis should be placed on individuals owning or controlling their own personal data, and retaining the ability to do with it as they please. In the light of the GDPR’s requirement for data portability, there are signs that this could become a reality, with a range of initiatives aimed at enabling individuals to control what happens to their data. For example, Sir Tim Berners-Lee is currently working on ‘Solid’, a proposed set of standards and tools, based on the idea of linked data, which would allow individuals on the internet to choose where to keep their personal data and how it should be used.¹⁰¹ The Hub of All Things (HAT) project, supported by the Engineering and Physical Sciences Research Council (EPSRC) and involving researchers from six British universities, seeks to help users take back control of their personal data by creating decentralised personal databases, controlled by individuals, allowing them to see all of their personal data in one place, see how it is being used, and sell or exchange it in return for money or benefits in kind.¹⁰² The Open Banking initiative, launched in January of this year, is also demonstrating how individual control of personal data, albeit only financial data for the time being, can work in practice.¹⁰³
79. Given the varied positions on data outlined above, we believe there is an urgent need for conceptual clarity on the subject of data if we are to prevent confusion in this area from hindering the development of AI. Datasets have properties which make them quite unlike most physical assets or resources in the world today. For example, datasets can be duplicated at near-zero cost, used in multiple ways by different people without diminishing their value, and their value often increases as they are combined with other datasets. Indeed, the question of how data can be accurately valued, and whether it can be treated as a form of property or type of asset, is an ongoing area of debate among economists, accountants, statisticians and other experts.¹⁰⁴

100 Written evidence from Dr Mike Lynch (AIC0005)

101 Solid, ‘What is Solid?’: <https://solid.mit.edu/> [accessed 5 February 2018]

102 Hub of All Things, ‘Personal Data Economy’: <https://hubofallthings.com/c/pde> [accessed 5 February 2018]

103 Written evidence from Competition and Markets Authority (AIC0245)

104 For recent policy-orientated discussions over the treatment of data as a form of intangible asset, and the need for robust methods of valuation, see Royal Academy of Engineering and the Institute of Engineering and Technology, *Connecting data: Driving productivity and innovation* (November 2015): <https://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity> [accessed 20 February 2018]; Professor Sir Charles Bean, *Independent Review of UK Economic Statistics* (March 2016): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf [accessed 20 February 2018]; Royal Society and British Academy, *Data management and use: Governance in the 21st century* (June 2017): <https://royalsociety.org/~media/policy/projects/data-governance/data-management-governance.pdf> [accessed 20 February 2018]

Box 2: Open Banking

Open Banking refers to a series of reforms relating to the handling of financial information by banks. From 13 January 2018, UK-regulated banks have had to let their customers share their financial data with third parties (such as budgeting apps, or other banks). Banks are sharing customer data in the form of open APIs (application programming interfaces) which are used to provide integrated digital services. The intent of these reforms is to encourage competition and innovation, and to lead to more, and better, products for money management. Importantly, personal information can only be shared if the data subject (the person whose information it is) gives their express permission.

The Competition and Markets Authority said “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.”

Source: *Written evidence from Competition and Markets Authority (AIC0245)*

80. In its Industrial Strategy White Paper and its evidence to us, the Government announced the creation of a national Centre for Data Ethics and Innovation, the establishment of ‘data trusts’, which would facilitate the sharing of datasets between organisations and companies, and the updating of its 2016 data ethics framework. The first policy is a reflection of the Government’s manifesto commitment to “institute an expert Data Use and Ethics Commission to advise regulators and parliament on the nature of data use and how best to prevent its abuse”,¹⁰⁵ as well as the Royal Society’s recommendation for a national data stewardship body. The second policy, stemming from the Hall-Pesenti Review, is to establish ‘data trusts’, which will monitor and supervise the sharing of datasets between organisations and companies. The Review emphasised that these trusts would “not be a legal entity or institution, but rather a set of relationships underpinned by a repeatable framework, compliant with parties’ obligations, to share data in a fair, safe and equitable way”.¹⁰⁶
81. However, it is important to note that the Review’s conception of a ‘data trust’ appears to be different from the understanding of the term expressed in our written evidence. The Hall-Pesenti Review imagined data trusts as more co-operative, less top-down organisations, whereby individuals could opt-in to particular trusts, and have a say in their governance and how the personal data they provide is used.¹⁰⁷ It was also unclear from our session with the Rt Hon Matt Hancock MP (then Minister of State for Digital) and Lord Henley, Parliamentary Under-Secretary of State at the Department for Business, Energy and Industrial Strategy, whether the Government intends to address the central questions over the value of public and personal data, as they did not answer our questions on this subject.¹⁰⁸

105 The Conservative and Unionist Party, *Manifesto 2017: Forward Together, Our Plan for a Stronger Britain and a Prosperous Future* (May 2017), p 79: <https://s3-eu-west-1.amazonaws.com/2017-manifestos/Conservative+Manifesto+2017.pdf> [accessed 26 March 2018]

106 Professor Dame Wendy Hall and Dr Jérôme Pesenti, *Growing the artificial intelligence industry in the UK*, (15 October 2017), p 46: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf [accessed 31 January 2018]

107 For examples, see written evidence from University College London (UCL) (AIC0135) and Toby Phillips and Maciej Kuziemski (AIC0197)

108 Q 192 (Matt Hancock MP and Lord Henley) and Q 200 (Matt Hancock MP and Lord Henley)

82. **The Government plans to adopt the Hall-Pesenti Review recommendation that ‘data trusts’ be established to facilitate the ethical sharing of data between organisations. However, under the current proposals, individuals who have their personal data contained within these trusts would have no means by which they could make their views heard, or shape the decisions of these trusts. We therefore recommend that as data trusts are developed under the guidance of the Centre for Data Ethics and Innovation, provision should be made for the representation of people whose data is stored, whether this be via processes of regular consultation, personal data representatives, or other means.**
83. **Access to data is essential to the present surge in AI technology, and there are many arguments to be made for opening up data sources, especially in the public sector, in a fair and ethical way. Although a ‘one-size-fits-all’ approach to the handling of public sector data is not appropriate, many SMEs in particular are struggling to gain access to large, high-quality datasets, making it extremely difficult for them to compete with the large, mostly US-owned technology companies, who can purchase data more easily and are also large enough to generate their own. In many cases, public datasets, such as those held by the NHS, are more likely to contain data on more diverse populations than their private sector equivalents, and more control can be exercised before they are released.**
84. **We recommend that wherever possible and appropriate, and with regard to its potential commercial value, publicly-held data be made available to AI researchers and developers. In many cases, this will require Government departments and public organisations making a concerted effort to digitise their records in unified and compatible formats. When releasing this data, subject to appropriate anonymisation measures where necessary, data trusts will play an important role.**
85. **We support the approach taken by Transport for London, who have released their data through a single point of access, where the data is available subject to appropriate terms and conditions and with controls on privacy. The Centre for Data Ethics and Innovation should produce guidance on similar approaches. The Government Office for AI and GovTech Catalyst should work together to ensure that the data for which there is demand is made available in a responsible manner.**
86. **We acknowledge that open data cannot be the last word in making data more widely available and usable, and can often be too blunt an instrument for facilitating the sharing of more sensitive or valuable data. Legal and technical mechanisms for strengthening personal control over data, and preserving privacy, will become increasingly important as AI becomes more widespread through society. Mechanisms for enabling individual data portability, such as the Open Banking initiative, and data sharing concepts such as data trusts, will spur the creation of other innovative and context-appropriate tools, eventually forming a broad spectrum of options between total data openness and total data privacy.**

87. **We recommend that the Centre for Data Ethics and Innovation investigate the Open Banking model, and other data portability initiatives, as a matter of urgency, with a view to establishing similar standardised frameworks for the secure sharing of personal data beyond finance. They should also work to create, and incentivise the creation of, alternative tools and frameworks for data sharing, control and privacy for use in a wide variety of situations and contexts.**
88. **Increasingly, public sector data has value. It is important that public organisations are aware of the commercial potential of such data. We recommend that the Information Commissioner’s Office work closely with the Centre for Data Ethics and Innovation in the establishment of data trusts, and help to prepare advice and guidance for data controllers in the public sector to enable them to estimate the value of the data they hold, in order to make best use of it and negotiate fair and evidence-based agreements with private-sector partners. The values contained in this guidance could be based on precedents where public data has been made available and subsequently generated commercial value for public good. The Information Commissioner’s Office should have powers to review the terms of significant data supply agreements being contemplated by public bodies.**

Intelligible AI

89. Alongside consumer awareness, many witnesses highlighted the importance of making AI understandable to developers, users and regulators.
90. Several witnesses told us that while many AI systems are no more difficult to understand than conventional software, the newer generation of deep learning systems did present problems.¹⁰⁹ As discussed earlier, these systems rely on the feeding of information through many different ‘layers’ of processing in order to come to an answer or decision. The number and complexity of stages involved in these deep learning systems is often such that even their developers cannot always be sure which factors led a system to decide one thing over another. We received a great deal of evidence regarding the extent and nature of these so-called ‘black box’ systems.
91. The terminology used by our witnesses varied widely. Many used the term transparency, while others used interpretability or ‘explainability’, sometimes interchangeably. For simplicity, we will use ‘intelligibility’ to refer to the broader issue. Within this, from our evidence, we have identified two broad approaches to intelligibility—technical transparency and explainability—which we address in more detail below.
92. The extent to which it was considered that an AI system needed to be intelligible at all was very much dependent on the witnesses in question, and the purposes to which the AI system was to be put. A small minority of witnesses appeared to believe the necessity for intelligible AI systems was low, and that current systems, even those which make use of deep learning, more than meet these demands. Nvidia, for example, made the point that machine learning algorithms are often much shorter and simpler than conventional software coding, and are therefore in some respects easier to

¹⁰⁹ Written evidence from Medicines and Healthcare products Regulatory Agency ([AIC0134](#)); Braintree ([AIC0074](#)) and Dr Mike Lynch ([AIC0005](#))

understand and inspect.¹¹⁰ Dr Timothy Lanfear, Director of the EMEA Solution Architecture and Engineering team at Nvidia, told us:

“We are using systems every day that are of a level of complexity that we cannot absorb. Artificial intelligence is no different from that. It is also at a level of complexity that cannot be grasped as a whole. Nevertheless, what you can do is to break this down into pieces, find ways of testing it to check that it is doing the things you expect it to do and, if it is not, take some action”.¹¹¹

93. Other witnesses thought it was unrealistic to hold AI to a higher standard than human decision-making, which itself can often seem illogical or impenetrable.¹¹² Within AI development today, there are many different techniques being used, which all come with advantages and disadvantages, and trade-offs will be inevitable. For example, decision tree learning can be faster, easier to understand, and are usually less data hungry, whereas deep neural networks are often more data hungry, slower and much less easy to understand, but can be more accurate, given the right data. Some witnesses argued that if we restricted ourselves only to techniques which we could fully understand, we would also limit what could be accomplished with AI.¹¹³
94. The idea of restricting the use of unintelligible systems in certain important or safety-critical domains was also mentioned by many witnesses. Experts from the University of Edinburgh emphasised the human element, arguing that given the “completely unintelligible” nature of decisions made via deep learning, it would be more feasible to focus on outcomes, and “only license critical AI systems that satisfy a set of standardised tests, irrespective of the mechanism used by the AI component”.¹¹⁴ Examples of contexts in which a high degree of intelligibility was thought to be necessary included:
- Judicial and legal affairs;¹¹⁵
 - Healthcare;¹¹⁶
 - Certain kinds of financial products and services (for example, personal loans and insurance);¹¹⁷
 - Autonomous vehicles;¹¹⁸ and
 - Weapons systems.¹¹⁹

110 Written evidence from NVIDIA ([AIC0212](#))

111 [Q 44](#) (Dr Timothy Lanfear)

112 Written evidence from Deep Science Ventures ([AIC0167](#)); Dr Dan O’Hara, Professor Shaun Lawson, Dr Ben Kirman and Dr Conor Linehan ([AIC0127](#)) and Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams ([AIC0029](#))

113 Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams ([AIC0029](#)); Michael Veale ([AIC0065](#)); Five AI Ltd ([AIC0128](#)); University College London (UCL) ([AIC0135](#)) and Electronic Frontier Foundation ([AIC0199](#))

114 Written evidence from Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams ([AIC0029](#))

115 Written evidence from Joanna Goodman, Dr Paresh Kathrani, Dr Steven Cranfield, Chrissie Lightfoot and Michael Butterworth ([AIC0104](#)); Future Intelligence ([AIC0216](#)) and Ocado Group plc ([AIC0050](#))

116 Written evidence from Medicines and Healthcare products Regulatory Agency ([AIC0134](#)); PHG Foundation ([AIC0092](#)); SCAMPI Research Consortium, City, University of London ([AIC0060](#)) and Professor John Fox ([AIC0076](#))

117 Written evidence from Professor Michael Wooldridge ([AIC0174](#))

118 Written evidence from Five AI Ltd ([AIC0128](#)) and UK Computing Research Committee ([AIC0030](#))

119 Written evidence from Big Brother Watch ([AIC0154](#)) and Amnesty International ([AIC0180](#))

Technical transparency

95. One solution to the question of intelligibility is to try to increase the technical transparency of the system, so that experts can understand how an AI system has been put together. This might, for example, entail being able to access the source code of an AI system. However, this will not necessarily reveal why a particular system made a particular decision in a given situation.¹²⁰ Significantly, it does not include the data that was input into the system in a particular scenario, nor how that data was processed in order to arrive at a decision, and a system may behave very differently with different datasets.
96. We were also told that what transparency means can vary widely depending on the underlying purpose. Professor Chris Reed, Professor of Electronic Commerce Law, Queen Mary University of London, argued that:
- “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”.¹²¹
97. He argued that there are situations where a lack of transparency in advance may be acceptable, on the basis that the overall societal benefits are significant and any loss can be compensated for. Indeed, requiring explanations for all decisions in advance could limit innovation, as it would limit the capacity for a system to evolve and learn through its mistakes. On the other hand, he believed that transparency should be required where fundamental rights are put at risk.¹²²
98. What people need to know about how an AI system is operating, and why it is making particular decisions, will often be very different depending on whether they are developers, users, or regulators, auditors or investigators, for example.¹²³ In many cases, a full technical breakdown of a system will not be useful. Professor Spiegelhalter explained to us that in many cases, telling “everybody everything” could actively be unhelpful; instead, “the information [a user is receiving] has to be accessible; they have to get it, understand it to some extent and be able to critique it”.¹²⁴
99. **Based on the evidence we have received, we believe that achieving full technical transparency is difficult, and possibly even impossible, for certain kinds of AI systems in use today, and would in any case not be appropriate or helpful in many cases. However, there will be particular safety-critical scenarios where technical transparency is imperative, and regulators in those domains must have the power to mandate the use of more transparent forms of AI, even at the potential expense of power and accuracy.**

120 Q 22 (Professor Alan Winfield)

121 Written evidence from Professor Chris Reed (AIC0055)

122 *Ibid.*

123 Written evidence from IBM (AIC0160); Simul Systems Ltd (AIC0016); Imperial College London (AIC0214) and Professor Chris Reed (AIC0055)

124 Q 216 (Professor Sir David Spiegelhalter)

Explainability

100. An alternative approach is explainability, whereby AI systems are developed in such a way that they can explain the information and logic used to arrive at their decisions. We learned of a variety of technical solutions which are currently in development, which could help explain machine learning systems and their decisions. A variety of companies and organisations are currently working on explanation systems, which will help to consolidate and translate the processes and decisions made by machine learning algorithms into forms that are comprehensible to human operators.¹²⁵ A number of large technology companies, including Google, IBM and Microsoft, told us of their commitment to developing interpretable machine learning systems, which included Google’s ‘Glassbox’ framework for interpretable machine learning, and Microsoft’s development of best practices for intelligible AI systems.¹²⁶
101. In the evidence we received, it was frequently observed that Article 22 of the GDPR contains a ‘right to an explanation’ provision.¹²⁷ This provision gives an individual, when they have been subject to fully automated decision-making (and where the outcome has a significant impact on them), the right to ask for an explanation as to how that decision was reached, or to ask for a human to make the decision. This provision was considered relatively vague, and contains a number of limitations, and does not apply if a decision is based on explicit consent, if it does not produce a legal outcome of similar significant effect on the data subject, or if the process is only semi-automated.¹²⁸ Nevertheless, our witnesses generally thought that it provided added impetus to solve the problem of explainable AI systems.¹²⁹ It was also pointed out that with regards to some AI healthcare systems (which most witnesses believed should be explainable), there were likely to be additional legal requirements as a result of the EU In Vitro Diagnostic Devices Regulation (2017), which comes into force in May 2022.¹³⁰
102. Indeed, the GDPR already appears to have prompted action in the UK, with the Data Protection Bill going further still towards enshrining a ‘right to an explanation’ in UK law. When we asked Dr Wachter, one of Europe’s leading experts on the subject, what she thought of the developing Bill, she told us:
- “ ... it has been proposed that after a decision has been made, the individual has to be informed about the outcome, which is new and better than what the GDPR currently offers. It also states that data subjects should have the right to ask that the decision be reconsidered, or that the decision not be made by an algorithm. Both those things meet and exceed what is currently envisaged in the GDPR, and that is excellent”¹³¹

125 Written evidence from Royal Academy of Engineering ([AIC0140](#)) and Michael Veale ([AIC0065](#))

126 Written evidence from Google ([AIC0225](#)); IBM ([AIC0160](#)) and Microsoft ([AIC0149](#))

127 Written evidence from Future Intelligence ([AIC0216](#)); PricewaterhouseCoopers LLP ([AIC0162](#)); IBM ([AIC0160](#)); CognitionX ([AIC0170](#)); Big Brother Watch ([AIC0154](#)); Will Crosthwait ([AIC0094](#)) and Dr Maria Ioannidou ([AIC0082](#))

128 Written evidence from Article 19 ([AIC0129](#))

129 Written evidence from Article 19 ([AIC0129](#)); Information Commissioner’s Office ([AIC0132](#)); CBI ([AIC0114](#)) and Ocado Group plc ([AIC0050](#))

130 Written evidence from PHG Foundation ([AIC0092](#))

131 [Q 58](#) (Dr Sandra Wachter)

103. However, we were also concerned to hear that the automated decision safeguards in the Data Protection Bill still only apply if a decision is deemed to be “significant” and based “solely on automated processing”.¹³² As we discuss elsewhere in this report, many AI systems today and into the future are aiming to augment, rather than fully replace, human labour, and as Michael Veale, a researcher at UCL, told us, “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”.¹³³ Veale argued that the law should also cover systems where AI is only part of the final decision, as is the case in France’s Digital Republic Act 2016, and that there was a need to incentivise the development of explainable AI systems, without having to rely solely on regulation.¹³⁴
104. The style and complexity of explanations will need to vary based on the audience addressed and the context in which they are needed. For example, the owner of a self-driving car will want certain kinds of information when asking their vehicle to explain why it chose a particular route, while an accident investigator will want very different kinds of information, probably at far more granular levels of detail, when assessing why the same vehicle crashed. Nevertheless, we should insist that these systems are built into AI products before they are deployed and that certain standards are met.
105. **We believe that the development of intelligible AI systems is a fundamental necessity if AI is to become an integral and trusted tool in our society. Whether this takes the form of technical transparency, explainability, or indeed both, will depend on the context and the stakes involved, but in most cases we believe explainability will be a more useful approach for the citizen and the consumer. This approach is also reflected in new EU and UK legislation. We believe it is not acceptable to deploy any artificial intelligence system which could have a substantial impact on an individual’s life, unless it can generate a full and satisfactory explanation for the decisions it will take. In cases such as deep neural networks, where it is not yet possible to generate thorough explanations for the decisions that are made, this may mean delaying their deployment for particular uses until alternative solutions are found.**
106. **The Centre for Data Ethics and Innovation, in consultation with the Alan Turing Institute, the Institute of Electrical and Electronics Engineers, the British Standards Institute and other expert bodies, should produce guidance on the requirement for AI systems to be intelligible. The AI development sector should seek to adopt such guidance and to agree upon standards relevant to the sectors within which they work, under the auspices of the AI Council.**

132 [Data Protection Bill](#), clause 14 [Bill 153 (2017–19)]

133 Written evidence from Michael Veale ([AIC0065](#))

134 The Digital Republic Act 2016, which covers French public bodies, refers to decisions ‘taken on the basis of algorithmic processing’, rather than ‘solely on automated processing’.

Addressing prejudice

107. The current generation of AI systems, which have machine learning at their core, need to be taught how to spot patterns in data, and this is normally done by feeding them large bodies of data, commonly known as training datasets. These systems are designed to spot patterns, and if the data is unrepresentative, or the patterns reflect historical patterns of prejudice, then the decisions which they make may be unrepresentative or discriminatory as well. This can present problems when these systems are relied upon to make real world decisions. Within the AI community, this is commonly known as ‘bias’.
108. While the term ‘bias’ might at first glance appear straightforward, there are in fact a variety of subtle ways in which bias can creep into a system. Much of the data we deem to be useful is about human beings, and is collected by human beings, with all of the subjectivity that entails. As LexisNexis UK put it, “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”.¹³⁵ It is also important to be aware that bias can emerge when datasets inaccurately reflect society, but it can also emerge when datasets accurately reflect unfair aspects of society.
109. For example, an AI system trained to screen job applications will typically use datasets of previously successful and unsuccessful candidates, and will then attempt to determine particular shared characteristics between the two groups to determine who should be selected in future job searches. While the intention may be to ascertain those who will be capable of doing the job well, and fit within the company culture, past interviewers may have consciously or unconsciously weeded out candidates based on protected characteristics (such as age, sexual orientation, gender, or ethnicity) and socio-economic background, in a way which would be deeply unacceptable today.
110. While some witnesses referred to the issue of bias as something which simply needed to be removed from training data and the AI systems developed using them,¹³⁶ others pointed out that this was more complicated. Dr Ing. Konstantinos Karachalios, Managing Director, IEEE-Standards Association, told us:
- “You can never be neutral; it is us. This is projected in what we do. It is projected in our engineering systems and algorithms and the data that we are producing. The question is how these preferences can become explicit, because if it can become explicit it is accountable and you can deal with it. If it is presented as a fact, it is dangerous; it is a bias and it is hidden under the table and you do not see it. It is the difficulty of making implicit things explicit”.¹³⁷
111. Dr Ansgar Koene made a similar point, when he distinguished between ‘operationally-justified bias’, which “prioritizes certain items/people as part of performing the desired task of the algorithm, e.g. identifying frail individuals when assigning medical prioritisation”, and ‘non-operationally-

135 Written evidence from LexisNexis UK ([AIC0164](#))

136 [Q 80](#) (James Luke)

137 [Q 24](#) (Dr Ing Konstantinos Karachalios)

justified bias’, which is “not integral to being able to do the task, and is often unintended and its presence is unknown unless explicitly looked for”.¹³⁸ Fundamentally, the issues arise when we are unaware of the hidden prejudices and assumptions which underpin the data we use. Olivier Thereaux noted Maciej Ceglowski’s description of machine learning as “like money laundering for bias.” Thereaux said:

“We take bias, which in certain forms is what we call ‘culture’, put it in a black box and crystallise it for ever. That is where we have a problem. We have even more of a problem when we think that that black box has the truth and we follow it blindly, and we say, ‘The computer says no’. That is the problem”.¹³⁹

112. Several witnesses pointed out that the issue could not be easily confined to the datasets themselves, and in some cases “bias and discrimination may emerge only when an algorithm processes particular data”.¹⁴⁰ The Centre for the Future of Intelligence emphasised that “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)”.¹⁴¹ The difficulties in fixing these issues was illustrated recently when it emerged that Google has still not fixed its visual identification algorithms, which could not distinguish between gorillas and black people, nearly three years after the problem was first identified. Instead, Google has simply disabled the ability to search for gorillas in products such as Google Photos which use this feature.¹⁴²
113. The consequences of this are already starting to be felt. Several witnesses highlighted the growing use of AI within the US justice system, in particular the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system, developed by Northpointe, and used across several US states to assign risk ratings to defendants, which help to assist judges in sentences and parole decisions. Will Crosthwait, co-founder of AI start-up Kensai, highlighted investigations which found that the system commonly overestimated the recidivism risk of black defendants, and underestimated that of white defendants.¹⁴³ Big Brother Watch observed that here in the UK, Durham Police have started to investigate the use of similar AI systems for determining whether suspects should be kept in custody, and described this and other developments as a “very worrying trend, particularly when the technology is being trialled when its abilities are far from accurate”.¹⁴⁴ Evidence from Sheena Urwin, Head of Criminal Justice at Durham Constabulary, emphasised the considerable lengths that Durham Constabulary have taken to ensure their use of these tools is open, fair and ethical, in particular the development of their ‘ALGO-CARE’ framework for the ethical use of algorithms in policing.¹⁴⁵

138 Written evidence from Dr Ansgar Koene (AIC0208)

139 Q 74 (Olivier Thereaux)

140 Written evidence from Research Councils UK (AIC0142)

141 Written evidence from Leverhulme Centre for the Future of Intelligence (AIC0182)

142 Tom Simonite, ‘When it comes to gorillas, Google Photos remains blind’, *Wired* (11 January 2018): <https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/> [accessed 31 January 2018]

143 Written evidence from Will Crosthwait (AIC0094)

144 Written evidence from Big Brother Watch (AIC0154)

145 Written evidence from Marion Oswald and Sheena Urwin (AIC0068)

114. A number of solutions were offered in the evidence we received. The most immediate was the creation of more diverse datasets, which fairly reflect the societies and communities which AI systems are increasingly affecting. Kriti Sharma told us “we now have the ability to create [diverse datasets]. If data does not exist we need to work hard, we need to work together and focus on open datasets”.¹⁴⁶ Several witnesses highlighted the role the Government could play in this regard, by releasing open datasets which are representative of the entire population, and which would address some of the shortcomings of less demographically-diverse privately-held datasets.¹⁴⁷ Witnesses frequently mentioned that there are a variety of groups in society, such as the financially excluded and ethnic minorities, who suffer from ‘data poverty’, especially compared to more privileged groups in society, who are likely to generate more data about themselves through a plethora of smart devices and services.¹⁴⁸ Dr Bunz also recommended that the Government should create guidance for the creation of datasets, which could help to minimise bias.¹⁴⁹
115. The importance of ensuring that AI development is carried out by diverse workforces, who can identify issues with data and algorithm performance, was also emphasised—a point we return to below.¹⁵⁰ Dr Wachter and academics from the Centre for the Future of Intelligence also argued that a greater diversity of academic disciplines needed to be involved in this process. Dr Wachter observed that questions around bias and prejudice in society have long been within the remit of philosophers, social scientists, legal theorists and political scientists, and warned that “if we do not have an interdisciplinary approach to these questions, we are going to leave out very important issues”.¹⁵¹
116. Many witnesses told us of the need to actively seek out bias within AI systems, by testing datasets and how they operate within particular AI systems. Dr Wachter believed that in cases where a system is “inherently opaque and not understandable”, as with many deep learning systems, “auditing after the fact, auditing during data processing or inbuilt processes that could detect biases” should be considered, backed up by certification of some form.¹⁵² Dr Koene told us of work already moving in this direction, which aims to establish a ‘Standard on Algorithm Bias Considerations’.¹⁵³
117. However, Kriti Sharma noted that “interesting research work has been done, but it has not been commercialised ... [the area] needs more funding from government”.¹⁵⁴ She suggested that the Challenge Fund, or other bodies working on AI, robotics and automation needed to divert more attention to the issue.

146 [Q 80](#) (Kriti Sharma)

147 Written evidence from Dr Mercedes Bunz ([AIC0048](#)); University College London ([AIC0135](#)) and Center for Data Innovation ([AIC0043](#))

148 Written evidence from Center for Data Innovation ([AIC0043](#)) and Weightmans LLP ([AIC0080](#))

149 Written evidence from Dr Mercedes Bunz ([AIC0048](#))

150 Written evidence from CognitionX ([AIC0170](#))

151 [Q 57](#) (Dr Sandra Wachter)

152 [Q 62](#) (Dr Sandra Wachter)

153 This will aim to “provide certification oriented methodologies to identify and mitigate non-operationally-justified algorithm biases through: the use of benchmarking procedures, criteria for selecting bias validation data sets and guidelines for the communication of application domain limitations (using the algorithm for purposes beyond this scope invalidates the certification).” Written evidence from Dr Ansgar Koene ([AIC0208](#))

154 [Q 80](#) (Kriti Sharma)

118. The message we received was not entirely critical though. A number of witnesses emphasised that, if the right measures are taken, we also have an opportunity to better address long-standing prejudices and inequalities in our societies. Kriti Sharma explained:

“AI can help us fix some of the bias as well. Humans are biased; machines are not, unless we train them to be. AI can do a good job at detecting unconscious bias as well. For example, if feedback is given in performance reviews where different categories of people are treated differently, the machine will say, ‘That looks weird. Would you like to reconsider that?’”¹⁵⁵

119. **We are concerned that many of the datasets currently being used to train AI systems are poorly representative of the wider population, and AI systems which learn from this data may well make unfair decisions which reflect the wider prejudices of societies past and present. While many researchers, organisations and companies developing AI are aware of these issues, and are starting to take measures to address them, more needs to be done to ensure that data is truly representative of diverse populations, and does not further perpetuate societal inequalities.**
120. **Researchers and developers need a more developed understanding of these issues. In particular, they need to ensure that data is pre-processed to ensure it is balanced and representative wherever possible, that their teams are diverse and representative of wider society, and that the production of data engages all parts of society. Alongside questions of data bias, researchers and developers need to consider biases embedded in the algorithms themselves—human developers set the parameters for machine learning algorithms, and the choices they make will intrinsically reflect the developers’ beliefs, assumptions and prejudices. The main ways to address these kinds of biases are to ensure that developers are drawn from diverse gender, ethnic and socio-economic backgrounds, and are aware of, and adhere to, ethical codes of conduct.**
121. **We recommend that a specific challenge be established within the Industrial Strategy Challenge Fund to stimulate the creation of authoritative tools and systems for auditing and testing training datasets to ensure they are representative of diverse populations, and to ensure that when used to train AI systems they are unlikely to lead to prejudicial decisions. This challenge should be established immediately, and encourage applications by spring 2019. Industry must then be encouraged to deploy the tools which are developed and could, in time, be regulated to do so.**

Data monopolies

122. As previously discussed, access to, and control of, data is a crucial ingredient in the development of modern AI, and data network effects can quickly lead to certain companies building up proprietary dataset dominance in the market. These can be difficult for smaller competitors to match. These dominant companies—sometimes referred to as the ‘tech giants’ or increasingly ‘Big Tech’—are commonly identified as Alphabet (Google’s parent company),

Amazon, Apple, Facebook, and Microsoft (as well as occasionally IBM, and beyond the USA, Samsung, Alibaba and Tencent), and have built business models partially, or largely, focused on the “aggregation of data and provision of cloud services”.¹⁵⁶ They are large and diverse enough that they can generate their own massive and varied datasets—for example, all of the data provided to Google through web searches and the use of Android smartphones, or data on social connections and interests provided to Facebook—without relying on third parties, and use this to understand consumer behaviour to an unprecedented extent.

123. These datasets, sometimes evocatively described within the sector as ‘data moats’ to signify the unassailable commercial positions they are meant to create, gives these companies a key advantage in the development of AI systems. This form of market dominance is increasingly referred to as a ‘data monopoly’.¹⁵⁷
124. We asked our witnesses if they had concerns about the dominance of these companies. Innovate UK said these organisations have “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”.¹⁵⁸ The Information Commissioner described the dominance of technology giants as “a vexing problem”.¹⁵⁹

Box 3: The Competition and Markets Authority

The Competition and Markets Authority (CMA) is a non-ministerial department which exists to promote competition for the benefit of consumers, both within and outside the UK. Their aim is to make markets work well for consumers, businesses and the economy. The CMA was established in April 2014 as a result of a merger of the Competition Commission and the Office of Fair Trading.

Source: Competition and Markets Authority, ‘About us’: <https://www.gov.uk/government/organisations/competition-and-markets-authority/about> [accessed 7 February 2018]

125. Some of our witnesses were less concerned by this dominance. The Center for Data Innovation, an international think tank which focuses on data, technology and public policy, said “there are no ‘data-based monopolies’ and the winner does not take all” as “data is non-rivalrous: customers who give their personal data to one company can provide it again to another”.¹⁶⁰ IBM told us that “any concerns about dominance can be addressed through competition law on an ex-post basis rather than a-priori regulation of AI” and that “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”.¹⁶¹ Capco said “there is no natural way to address these monopolies and perhaps they should not be addressed while they serve the consumer and society” but regulators should be vigilant in making sure that these companies do not inhibit competition.¹⁶²

156 Written evidence from Royal Academy of Engineering (AIC0140)

157 Though technically, given the presence of several large companies, it more closely resembles an oligopoly, rather than a monopoly.

158 Written evidence from Innovate UK (AIC0220)

159 Q 59 (Elizabeth Denham)

160 Written evidence from the Center for Data Innovation (AIC0043)

161 Written evidence from IBM (AIC0160)

162 Written evidence from Capco (AIC0071)

126. Digital Catapult told us these large companies do not “engage in typical monopoly behaviour such as stealing market share by selling goods below the cost of production” and suggested that their success has benefited consumers: “few people would be happy to give up using Amazon’s one day delivery or Google search”.¹⁶³ Digital Catapult did also, however, state that “data is becoming even more valuable” and the large companies must not restrict SME access to data or prevent the growth of an AI developer sector.¹⁶⁴
127. Other witnesses felt it was too insurmountable a problem: “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”.¹⁶⁵ Many of our witnesses, however, believed there to be serious issues with what they believed to be the monopoly-building activities of large, mostly US-headquartered, technology companies, and identified ways to address the issue.
128. The Information Commissioner told us “people are becoming more and more concerned about information monopolies and large platforms that offer many different services and collect and link all kinds of personal data”.¹⁶⁶ Other witnesses said that people were not aware enough of how their personal data was being used.¹⁶⁷ Professor Kathleen Richardson and Ms Nika Mahnič, from the Campaign Against Sex Robots, described the giving of personal information for access to digital services as a “Faustian pact”, and that people need to be more aware of the implications.¹⁶⁸ A concern over how one’s personal data is used is a fair response when it comes to the development and deployment of artificial intelligence. Such concerns can also be best addressed by individuals having greater control over their data, as we have discussed above.
129. **While we welcome the investments made by large overseas technology companies in the UK economy, and the benefits they bring, the increasing consolidation of power and influence by a select few risks damaging the continuation, and development, of the UK’s thriving home-grown AI start-up sector. The monopolisation of data demonstrates the need for strong ethical, data protection and competition frameworks in the UK, and for continued vigilance from the regulators. We urge the Government, and the Competition and Markets Authority, to review proactively the use and potential monopolisation of data by the big technology companies operating in the UK.**

163 Written evidence from Digital Catapult ([AIC0175](#))

164 *Ibid.*

165 Written evidence from Advanced Marine Innovation Technology Subsea Ltd ([AIC0038](#))

166 [Q 59](#) (Elizabeth Denham)

167 See written evidence from Department of Computer Science University of Liverpool ([AIC0192](#)); BBC ([AIC0204](#)) and Future Intelligence ([AIC0216](#))

168 Written evidence from Professor Kathleen Richardson and Ms Nika Mahnič ([AIC0200](#))

CHAPTER 4: DEVELOPING ARTIFICIAL INTELLIGENCE

130. The UK is currently one of the best countries in the world for researchers and businesses developing artificial intelligence. This cannot, however, be taken for granted. This chapter focuses on the challenges the UK may face in maintaining its position as a world leader, and offers solutions as to how we, as a nation, can ensure that the environment for businesses developing AI continues in good health, and that we have access to the best global talent.

Investment in AI development

131. A number of countries are currently investing significant sums in AI research and development, with varying degrees of state support and co-ordination. Estimating the size of these investments across the public and private sectors is difficult, but according to Goldman Sachs, between the first quarter of 2012 and the second quarter of 2016, the United States invested approximately \$18.2 billion in AI, compared with \$2.6 billion by China, and \$850 million in the UK, the third highest investment by a country in this period.¹⁶⁹ While the current United States administration appears to have no national-level AI strategy, individual government departments are continuing to invest in AI, with the Department of Defense, for example, spending approximately \$2.5 billion on AI in 2017.¹⁷⁰ Meanwhile, China has explicitly committed itself to becoming a world leader in AI by 2030, and aims to have grown its AI ecosystem to \$150 billion by then.¹⁷¹
132. Given the disparities in available resources, the UK is unlikely to be able to rival the scale of investments made in the United States and China. Germany and Canada offer more similar comparisons. Of these two, Germany's approach is strongly influenced by its flagship Industrie 4.0 strategy for smart manufacturing. This strategy seeks to use AI to improve manufacturing processes, and to produce 'smart goods' with integrated AI, such as fridges and cars.¹⁷² As Professor Wolfgang Wahlster, CEO and Scientific Director of the German Research Center for AI (DFKI), told us, "this is quite different from the US approach, which is based more on internet services".¹⁷³ Meanwhile, the Pan-Canadian Artificial Intelligence Strategy, which we discuss in Chapter 9, is less focused on developing AI in any particular sector, but is making C\$125 million available to establish three new AI institutes across the country, and attract AI researchers to the country.¹⁷⁴

169 Goldman Sachs, *China's rise in artificial intelligence* (31 August 2017), p 6: <http://www2.caict.ac.cn/zscp/qqzkgz/ljyd/201709/P020170921309379565253.pdf> [accessed 23 February 2018]

170 Govini, *Department of Defense: artificial intelligence, big data and cloud taxonomy* (December 2017): <http://www.govini.com/dod-ai-big-data-and-cloud-taxonomy/> [accessed 7 March 2018]

171 Indeed, a number of commentators have pointed out that China's AI strategy appears to have been influenced by the Obama administration's AI strategy, published in late 2016, which has since become dormant under the Trump administration. Cade Metz, 'As China marches forward on AI, the White House is silent', *New York Times* (12 February 2018): <https://www.nytimes.com/2018/02/12/technology/china-trump-artificial-intelligence.html> [accessed 13 February 2018]

172 Q 164 (Professor Wolfgang Wahlster)

173 *Ibid.*

174 CIFAR, 'Pan-Canadian artificial intelligence strategy overview' (30 March 2017): <https://www.cifar.ca/assets/pan-canadian-artificial-intelligence-strategy-overview/> [accessed 21 February 2018]

133. Over the course of our inquiry, we were often told of the vitality of the UK AI development sector. The Hall-Pesenti Review recognised this, stating: “The UK has AI companies that are seen as some of the world’s most innovative, in an ecosystem that includes large corporate users of AI, providers large and small, business customers for AI services, and research experts”.¹⁷⁵ Investment, as is the case with Canada and unlike in Germany, is not focused in any particular area, and the Government has, for the most part, declined to direct research priorities, with Lord Henley characterising this approach as “[letting] a hundred flowers bloom”.¹⁷⁶
134. David Kelnar, Head of Research at MMC Ventures, told us “the number of AI start-ups founded annually in the UK has doubled since 2014, and since then, on average, a new AI start-up has been founded every five days in the UK”.¹⁷⁷ The Hall-Pesenti Review estimated that there were more than 200 start-ups and SMEs developing AI products in the UK, as of October 2017.¹⁷⁸ Dr Marko Balabanovic, Chief Technology Officer, Digital Catapult, told us that Digital Catapult had found that there were around 600 AI start-ups in the UK out of a total of 1,200 in Europe, putting the UK in a good position.¹⁷⁹
135. **The UK AI development sector has flourished largely without attempts by the Government to determine its shape or direction. This has resulted in a flexible and innovative grassroots start-up culture, which is well positioned to take advantage of the unpredictable opportunities that could be afforded by AI. The investment environment for AI businesses must be able to cope with this uncertainty, and be willing to take the risks required to seize the chances AI offers.**

Box 4: Start-ups and SMEs

‘Start-up’ is a term usually used to refer to businesses which have recently established themselves. They are typically small, financed and operated by their founders, and usually attempt to offer innovative solutions to a shared problem. There is, however, no agreed definition of what a ‘start-up’ is, and companies as large as Uber and WhatsApp often self-define as start-ups.

Medium and small employers are often referred to as small and medium sized enterprises (or SMEs). A business is normally considered to be an SME if it employs between 10 and 249 staff.

Source: Organisation for Economic Co-operation and Development (OECD), ‘Glossary of Statistical Terms’: <https://stats.oecd.org/glossary/detail.asp?ID=3123> [accessed 7 February 2018]

136. Eileen Burbidge MBE, Chair of Tech City UK, a Partner at Passion Capital, and a Treasury special envoy for fintech, told us there was a healthy appetite for investing in start-ups in the UK, and that “that there is no shortage of capital at every stage of a life cycle” in Britain.¹⁸⁰

175 *Growing the artificial intelligence industry in the UK*, p 23

176 [Q 191](#) (Lord Henley)

177 [Q 48](#) (David Kelnar)

178 *Growing the artificial intelligence industry in the UK*, p 24

179 [Q 42](#) (Dr Marko Balabanovic)

180 [Q 48](#) (Eileen Burbidge)

137. Other witnesses disagreed, and told us of the difficulties that UK investors face in competing with the largest American technology companies and investors based in Silicon Valley. Libby Kinsey, co-founder of Project Juno, told us: “European investors tend to have smaller funds and less appetite for risk than US investors”.¹⁸¹ Dr David Barber, Reader in Computational Statistics and Machine Learning, UCL, echoed this: “there are a lot of successful tech start-ups which came out of the UK which in the end became successful only when they went to Silicon Valley for investment”.¹⁸²
138. It was clear from discussions with technology companies on our visit to Cambridge and in the course of our roundtable at techUK (see Appendices 6 and 8) that the appetite for investing in start-ups did exist in the UK, but that the lifecycle and scaling finance was less available. The Royal Society stated “the recent acquisitions of DeepMind, VocalIQ, Swiftkey, and Magic Pony, by Google, Apple, Microsoft, and Twitter respectively, point to the success of UK start-ups in this sector”.¹⁸³ The Royal Society also, however, voiced disquiet at these acquisitions of UK-based start-ups by foreign-owned companies, stating “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”.¹⁸⁴ The reported cost of these acquisitions ranges from \$50–100 million (for VocalIQ) up to £650 million for DeepMind. Professor Michael Wooldridge, Head of the Department of Computer Science, University of Oxford, acknowledged that “there is an incredibly vibrant start-up culture in London”.¹⁸⁵ He also warned that “it is fragile and needs to be nurtured”.¹⁸⁶
139. James Wise, a Partner at Balderton Capital (UK) LLP, highlighted the specific challenges for AI-focused companies:
- “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”.¹⁸⁷
140. Many of the problems are shared with the wider technology sector in the UK, but the potential of artificial intelligence for the UK’s economy means the challenge in scaling up is even more problematic. Recent Government announcements and the Green Investment Bank (GIB) model offer lessons as to how the Government can use its influence to improve the environment for AI start-ups, and enable them to scale up.

181 [Q 49](#) (Libby Kinsey)

182 [Q 42](#) (Dr David Barber)

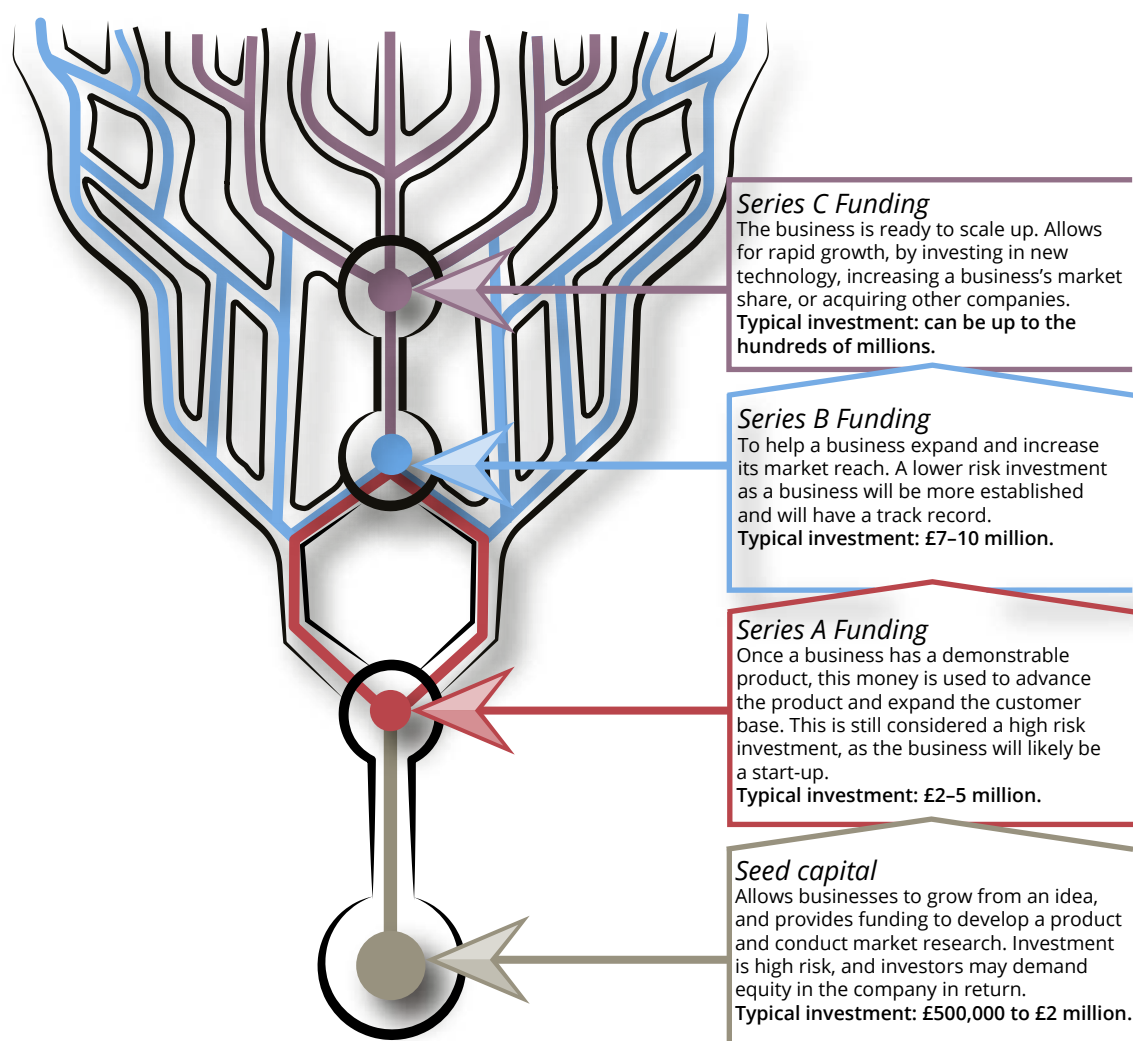
183 Written evidence from Royal Society ([AIC0168](#))

184 *Ibid.*

185 [Q 3](#) (Professor Michael Wooldridge)

186 *Ibid.*

187 Written evidence from Balderton Capital (UK) LLP ([AIC0232](#))

Figure 3: Investment rounds

141. The GIB was established by the Coalition Government in October 2012 to finance the green economy—an economy that aims to support sustainable development without degrading the environment—and “to accelerate private sector investment, with an initial remit to focus on relatively high-risk projects that are otherwise likely to proceed slowly or not at all”.¹⁸⁸ By March 2017 the GIB had invested in 100 projects, committing up to £3.4 billion of its own capital, and had attracted £8.6 billion of private capital. In August 2017, the GIB was sold to Macquarie, a global investment banking and diversified financial services group, for £1.6 billion.
142. The National Audit Office (NAO) reported in December 2017 on the effectiveness of the GIB, finding that “it quickly stimulated investment in the green economy”, in part because its structure as a public company gave it the freedom to pursue its objectives and intentionally constrained its investment activities. The NAO also said that the GIB had invested in, and attracted private capital to, each of its approved sectors. The NAO reported that the responsible Department (then the Department for Business, Innovation and Skills, now the Department for Business, Energy and Industrial Strategy)

had not established clear criteria or evidence to judge whether the GIB was achieving its intended green impact. The NAO was also critical of the way in which the sale of the GIB was handled.¹⁸⁹

143. This example shows that concerted policy interventions to incentivise private investment for the public good can work, if such policy interventions are committed to.
144. The Autumn Budget 2017 included a series of policy announcements to encourage the growth of innovative firms in the UK. Two of the most relevant were the establishment of a £2.5 billion Investment Fund within the British Business Bank, and proposed changes to the Enterprise Investment Scheme (EIS) and the Venture Capital Trust scheme (VCT). The Rt Hon Philip Hammond MP, Chancellor of the Exchequer, described these measures in the House of Commons as “an action plan to unlock over £20 billion of new investment in UK knowledge-intensive, scale-up businesses”.¹⁹⁰
145. The British Business Bank is a Government-owned economic development bank, formed in 2014, with the aim of increasing the supply of credit to SMEs. The Government announced in its November 2017 Budget that the British Business Bank would be responsible for a new £2.5 billion Investment Fund. The British Business Bank said this fund could “unlock up to £13 billion of finance to support UK smaller businesses looking to scale-up and realise their growth potential”.¹⁹¹ Matt Hancock MP highlighted the value of the British Business Bank: “Approximately 40% of capital at some stages in the market is backed in some way by the British Business Bank, so we [the Government] should not underestimate the role that we are playing in this space”.¹⁹²
146. The EIS is a series of tax reliefs designed to encourage investments in small companies operating in the UK, and has been in operation since 1994. The VCT is another long-standing scheme, designed to encourage people to invest indirectly in a range of unlisted, smaller, higher-risk trading companies, by investing through a VCT instead of directly. The proposed changes to the EIS include doubling the annual allowance for people investing in knowledge-intensive companies (from £1 million to £2 million), and increasing the annual investment such companies receive through the EIS and the VCT (from £5 million to £10 million).¹⁹³ The EIS will also be altered to focus less on ‘low-risk’ businesses.
147. Finally, one other significant incentive for companies looking to invest in AI R&D are the R&D tax credit schemes. There are two main types of tax relief which companies may be eligible: SME R&D relief, aimed at companies with less than 500 staff and a turnover of less than €100 million or a balance sheet total of less than €86 million, and Research and Development Expenditure

189 National Audit Office, *The Green Investment Bank* (12 December 2017): <https://www.nao.org.uk/wp-content/uploads/2017/12/The-Green-Investment-Bank.pdf> [accessed 11 January 2018]

190 HC Deb, 22 November 2017, col 1049

191 British Business Bank, *Budget 2017* (23 November 2017): <https://british-business-bank.co.uk/budget-2017/> [accessed 12 January 2018]

192 Q 194 (Matt Hancock MP)

193 HM Treasury, *Autumn Budget 2017* (November 2017), p 49: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/661480/autumn_budget_2017_web.pdf [accessed 17 January 2018]

Credit (RDEC), for larger companies.¹⁹⁴ To be eligible for R&D relief, a company must show how a project:

- looked for an advance in science and technology which could be applicable within the wider field;
- had to overcome uncertainty, and how they did this; and
- could not be easily worked out by a professional in the field.

Under the SME tax relief scheme, a company is allowed to deduct an extra 130% of their qualifying costs from their yearly profit, as well as the normal 100% deduction, to make a total 230% deduction. If the company is loss making, the tax credit can be worth up to 14.5% of the surrenderable loss. Under the RDEC and larger company R&D schemes, a tax credit worth 11% (soon to rise to 12%) of qualifying R&D expenditure may be claimed.¹⁹⁵

148. Relatively few of our witnesses mentioned R&D tax credits to us, though at our roundtable event with techUK, some attendees believed the current system unduly benefited larger companies and was not assisting SMEs as much as it could be.¹⁹⁶ In 2015–16, 21,865 of the 26,255 claims for R&D tax relief came from SMEs, and the number of SMEs claiming rose by 22% on the previous year.¹⁹⁷ Within the Information and Communication sub-category, the single category most directly applicable to AI development,¹⁹⁸ 5,805 claims were made under the SME R&D scheme, worth £385 million, compared with £145 million, across 355 claims, under larger company schemes.¹⁹⁹
149. In total, SMEs have claimed more in R&D tax relief than larger companies. However, 80% of the £22.9 billion in the total qualifying R&D expenditure used to claim R&D tax relief was by companies claiming under large company schemes. This suggests that a sizeable majority of the research and development being incentivised by tax relief schemes is actually being conducted by large companies rather than SMEs.²⁰⁰ Some business leaders have accordingly been critical of the decision to increase RDEC tax relief from 11 to 12% in the 2017 Autumn Budget, without any equivalent increase in the SME scheme. The Institute for Public Policy Research (IPPR) has argued that R&D tax credits and similar schemes for larger companies, which cost £1.8–1.9 billion annually, should be scrapped entirely and

194 HM Revenue & Customs, ‘Guidance: Research and development (R&D) tax reliefs’ (14 August 2017): <https://www.gov.uk/guidance/corporation-tax-research-and-development-rd-relief#types-of-rd-relief> [accessed 14 February 2018]

195 *Ibid.*

196 Written evidence from Bikal (AIC0052), BSA The Software Alliance (AIC0153) and techUK (AIC0203)

197 In addition, 1,770 SMEs also claimed for research under the RDEC scheme and large company R&D scheme, which is permissible if a larger company sub-contracts research to an SME. HM Revenue & Customs, *Research and Development Tax Credits Statistics* (September 2017): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/644599/2017_RD_publication_commentary_final.pdf [accessed 22 February 2018]

198 It should be noted that, due to the far-reaching nature of AI technology, in many cases it could also fall into other sub-categories of R&D.

199 Additionally, a total of £20 million was claimed by SMEs under larger company tax relief schemes.

200 SMEs receive more in tax relief, even though much more qualifying R&D is conducted by large companies, because incentives for each individual small company are substantially more generous than under the large company schemes. HM Revenue & Customs, *Research and Development Tax Credits Statistics* (September 2017): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/644599/2017_RD_publication_commentary_final.pdf [accessed 22 February 2018]

redirected towards SMEs.²⁰¹ Additionally, many AI start-ups are likely to be excluded from claiming under existing rules, as they are experimenting with the application of established AI techniques and methods to new sectors, which is explicitly not covered by R&D tax relief.²⁰²

150. **We welcome the changes announced in the Autumn Budget 2017 to the Enterprise Investment and Venture Capital Trust schemes which encourage innovative growth, and we believe they should help to boost investment in UK-based AI companies. The challenge for start-ups in the UK is the lack of investment available with which to scale up their business.**
151. **To ensure that AI start-ups in the United Kingdom have the opportunity to scale up, without having to look for off-shore investment, we recommend that a proportion of the £2.5 billion investment fund at the British Business Bank, announced in the Autumn Budget 2017, be reserved as an AI growth fund for SMEs with a substantive AI component, and be specifically targeted at enabling such companies to scale up. Further, the Government should consult on the need to improve access to funding within the UK for SMEs with a substantive AI component looking to scale their business.**
152. **To guarantee that companies developing AI can continue to thrive in the UK, we recommend that the Government review the existing incentives for businesses operating in the UK who are working on artificial intelligence products, and ensure that they are adequate, properly promoted to companies, and designed to assist SMEs wherever possible.**

Turning academic research into commercial potential

153. The UK has a proven record of producing world-class academic research at globally renowned universities. It has, however, struggled to produce the businesses and commercial success which could flow from this. In addition to this widely recognised problem, artificial intelligence presents its own challenges, in particular the pace and intellectual property involved in AI research and development.
154. This was a problem identified in the Hall-Pesenti Review: “a key component that drives the creation (and success) of new businesses in AI is the ability and capacity for ideas and technologies to spin out of the university network, or be licensed from it, and be commercialised”.²⁰³ The Review identified the potential complexity of spin-out practices, and the differing approaches taken by universities make it all the harder for researchers to succeed in commercialising their research. Recommendation 11 of the Review was that “universities should use clear, accessible and where possible common policies and practices for licensing IP and forming spin-out companies”.²⁰⁴

201 Zen Terrenlonge, ‘Autumn budget 2017: R&D tax credits and investment to propel UK into the future’, *Real Business* (22 November 2017): <https://realbusiness.co.uk/accounts-and-tax/2017/11/22/autumn-budget-2017-rd-tax-credits-investment-propel-uk-future/> [accessed 12 February 2018]; IPPR, *Industrial strategy: Steering structural change in the UK economy* (November 2017): https://www.ippr.org/files/2017-11/1511445722_industrial-strategy-cej-november17.pdf [accessed 7 March 2018]

202 HM Revenue & Customs, ‘Guidance: Research and development (R&D) tax reliefs’ (14 August 2017): <https://www.gov.uk/guidance/corporation-tax-research-and-development-rd-relief#types-of-rd-relief> [accessed 14 February 2018]

203 *Growing the artificial intelligence industry in the UK*, p 63

204 *Growing the artificial intelligence industry in the UK*, p 64

Box 5: What is a spin-out?

A spin-out company is not dissimilar to a start-up, but with the crucial difference that it will often have a minority shareholder, such as a higher education institute, alongside the founders as owners. Spin-outs offer a mechanism for universities to benefit from the research of its staff when they look to apply it commercially.

155. Our witnesses shared these concerns. They told us of the challenges faced by artificial intelligence researchers seeking to commercialise their work. The Royal Society told us:

“ ... this standard model [for typical university spin-out companies] may fit less well for machine learning spin-outs. There may not be any IP [intellectual property] 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”.²⁰⁵

156. David Kelnar agreed that the existing model for spin-out companies was a challenge for AI researchers, and said:

“Universities, typically, seek quite substantial ownership stakes in spin-outs in return for assets, such as patents, the substantial support they offer and the expectation of significant dilution of ownership that will occur over time due to the spinout’s large capital requirements. In the era of AI, though, researchers’ primary assets are more likely to be a little different—it is more a case of their expertise and capability rather than those existing assets”.²⁰⁶

157. Kelnar also identified that limited access to commercial experience and advice was also an issue.²⁰⁷ He told us that this was because there were “not very many commercially experienced AI leaders”.²⁰⁸

158. Witnesses highlighted the work of Imperial Innovations, a subsidiary of IP Group plc (a developer of intellectual property-based businesses, which works exclusively with Imperial College London).²⁰⁹ Imperial Innovations launched *Founder’s Choice* as a pilot programme for 18 months in August 2017. The aim of the programme was to help reduce the equity share requirements of Imperial Innovations, based on the changing nature of support required for, and by, spin-outs. While the outcome of the programme awaits to be seen, the initiative seems promising.

159. **The UK has an excellent track record of academic research in the field of artificial intelligence, but there is a long-standing issue with converting such research into commercially viable products.**

160. **To address this we welcome, and strongly endorse, the recommendation of the Hall-Pesenti Review, which stated “universities should use clear, accessible and where possible common policies and practices for licensing IP and forming spin-out companies”. We recommend that the Alan Turing Institute, as the National Centre for AI Research, should develop this concept into concrete policy advice for universities in the UK, looking to examples from other fields and from other nations, to help start to address this long-standing problem.**

205 Written evidence from Royal Society ([AIC0168](#))

206 [Q 50](#) (David Kelnar)

207 *Ibid.*

208 *Ibid.*

209 [Q 50](#) (David Kelnar) and written evidence from Balderton Capital (UK) LLP ([AIC0232](#))

Improving access to skilled AI developers

161. One of the most pressing roadblocks we heard about was the substantial shortfall in skilled workers available to the AI development sector in the UK. Almost all the companies and organisations active in AI development from whom we received evidence complained that developers with advanced knowledge of machine learning, particularly at the PhD and master's degree levels, were difficult to find, and expensive to hire. Balderton Capital told us that “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”.²¹⁰ The Royal Society also argued that “additional resources to increase this talent pool are critically needed”, with particular emphasis on increased provision for the training of PhD students in machine learning.²¹¹
162. At our roundtable event with SMEs, hosted by techUK, we also heard that a number of companies had taken it upon themselves to fund PhD students in machine learning. Drawing in PhD funding from the private sector is indeed encouraged by the research councils, as with the EPSRC's Doctoral Training Partnership (DTP) scheme, whereby the cost of funding a PhD is split between the research council and a private sponsor.²¹²
163. We were told how the high private sector demand for machine learning expertise risked eroding training pipelines, as the academics needed to train the next generation of talent were being attracted away from universities into private companies. The Future of Humanity Institute at the University of Oxford warned that high 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”.²¹³ They suggested that lessons might be learnt from “other domains, such as finance and law, where competition for talent with the private sector has been fierce”, and universities could “consider novel initiatives such as special authority for a department to pay higher than usual salaries”.²¹⁴
164. It was pointed out that while PhDs in machine learning and AI more widely were important, there was also a need for shorter postgraduate qualifications, such as master's degrees. The development of new machine learning platforms and tools, such as the open source TensorFlow from Google, means that the level of skill needed to deploy AI in a variety of circumstances, and to facilitate the adoption of AI-enabled services by companies outside the AI development sector, is steadily decreasing.
165. The Royal Academy of Engineering highlighted that there was 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”.²¹⁵ Furthermore, Research Councils UK stated, “a wider range [of skills] will be needed in the AI workforce as it increasingly overlaps

210 Written evidence by Balderton Capital (UK) LLP ([AIC0232](#))

211 Written evidence from the Royal Society ([AIC0168](#))

212 EPSRC, ‘Doctoral Training Partnership’: <https://www.epsrc.ac.uk/skills/students/dta/> [accessed 10 January 2018]

213 Written evidence from Future of Humanity Institute ([AIC0103](#))

214 *Ibid.*

215 Written evidence from Royal Academy of Engineering ([AIC0140](#))

with ethics and social sciences. For example economists are needed for the development of fintech systems, [and] linguists for the development of language processing systems”.²¹⁶

166. In the Hall-Pesenti Review, it was recommended that the Government and universities should create, at a minimum, an additional 200 PhD places dedicated to AI at leading universities, and develop new master’s level courses in AI, in collaboration with industry, with an initial cohort of 300 students. These recommendations have subsequently been adopted as Government policy, with an announcement that an additional 200 PhD places in AI-related subjects would be funded per year by 2020–22, and plans to work with universities and businesses to develop an industry-funded master’s programme in AI.²¹⁷
167. When we asked Dr Pesenti, co-chair of the Hall-Pesenti Review, for the rationale behind these numbers, he explained that there had been considerable discussion on this point with civil servants, and noted that they had ultimately been revised down to ensure the recommendation’s sustainability:

“If you look at the demand right now, it needs to be counted in the thousands, quickly, in the next decade for sure. You cannot get there tomorrow because people are not able to be trained. You need to have faculty and fellows, which we also recommended in the review. There was this question: should you put the big number first or should you start with 300? There, we got a lot of back-and-forth”.²¹⁸

After further questioning, he suggested it was important to think in terms of “tens of thousands” of PhD places within the next decade, but re-emphasised the importance of building up to this number in a sustainable way.²¹⁹

168. **We welcome the expanded public funding for PhD places in AI and machine learning, as well as the announcement that an industry-funded master’s degree programme is to be developed. We do believe that more needs to be done to ensure that the UK has the pipeline of skills it requires to maintain its position as one of the best countries in the world for AI research.**
169. **We recommend that the funding for PhD places in AI and machine learning be further expanded, with the financial burden shared equally between the public and private sector through a PhD matching scheme. We believe that the Doctoral Training Partnership scheme and other schemes where costs are shared between the private sector, universities and research councils should be examined, and the number of industry-sponsored PhDs increased.**
170. **We further recommend that short (3–6 months) post-graduate conversion courses be developed by the Alan Turing Institute, in conjunction with the AI Council, to reflect the needs of the AI development sector. Such courses should be suitable for individuals in other academic disciplines looking to transfer into AI development and design or to have a grounding in the application of AI in their discipline. These should be designed so as to enable anyone to retrain at any stage of their working lives.**

216 Written evidence from Research Councils UK ([AIC0142](#))

217 *Industrial Strategy: Building a Britain fit for the future*, p 39

218 [Q 203](#) (Dr Jérôme Pesenti)

219 [Q 212](#) (Dr Jérôme Pesenti)

Diversity of talent

171. Our witnesses raised questions over the diversity of those working in AI development. In the early decades of the computer industry there was once a significant proportion of female workers. Unfortunately this is no longer the case, and some of our witnesses spoke of the need to “democratise AI”, and address what Bill Gates has described as the “sea of dudes” problem, with mainly male attendees at AI conferences.²²⁰ PwC told us of research they had conducted, which found that only 27% of female students they surveyed said they would consider a career in technology, compared to 61% of males, and that only 3% of females said it would be their first choice. 78% of students surveyed could not name a famous woman working in technology, compared to two thirds that could name a famous man working in technology.²²¹ As a consequence, according to Liberty, only 7% of students taking the computer science A-Level, and 17% of those working in technology in the UK, are female.²²²
172. Ensuring that those from low-income households and disadvantaged socio-economic backgrounds can still participate in the development and adoption of AI was also raised as part of wider efforts to facilitate social mobility.²²³ During our meeting with representatives of UK AI start-ups, we heard some scepticism regarding the inflexibility of the apprenticeship levy with regards to the AI development sector. Paul Clarke, Chief Technology Officer at Ocado, told us that the apprenticeship levy had “carved a hole in the available budget that companies, including ours, have to spend on that continual learning”, and urged that it be converted into a less ring-fenced “training levy”.²²⁴ However, at least one company in attendance at the techUK roundtable told us of their success using apprentices in their company. PwC also informed us of their recently announced technology degree apprenticeship scheme, beginning 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.²²⁵
173. Gender, ethnic and socio-economic diversity are important for a variety of reasons. Careers in AI are well remunerated and an area of rapid growth, and the dominance of these positions by already privileged groups in society is likely to exacerbate existing inequalities further. But this lack of diversity also has a significant impact on the way that AI systems are designed and developed. If we are to ensure that these systems, which are exerting growing influence over our lives and societies, serve us fairly rather than perpetuate and exacerbate prejudice and inequality, it is important to ensure that all groups in society are participating in their development. As CognitionX put it, “one of the reliable ways we know we can mitigate [the problem of bias and discrimination] is to have more diverse development teams in terms of specialisms, identities and experience”.²²⁶ Companies are making efforts to address this issue, and we are aware that many issues need to be tackled within primary and secondary education, which we address later in this report.²²⁷ Nevertheless, we believe there are still measures the Government can take to address this problem in the short term.

220 Written evidence from Dr Huma Shah and Professor Kevin Warwick ([AIC0066](#))

221 Written evidence from PricewaterhouseCoopers LLP ([AIC0162](#))

222 Written evidence from Liberty ([AIC0181](#))

223 Written evidence from Vishal Wilde ([AIC0004](#))

224 [Q 107](#) (Paul Clarke)

225 Written evidence from PricewaterhouseCoopers LLP ([AIC0162](#))

226 Written evidence from CognitionX ([AIC0170](#))

227 Google ([AIC0225](#)) and PricewaterhouseCoopers LLP ([AIC0162](#))

174. **We recommend that the Government ensures that publically-funded PhDs in AI and machine learning are made available to a diverse population, more representative of wider society. To achieve this, we call for the Alan Turing Institute and Government Office for AI to devise mechanisms to attract more female and ethnic minority students from academic disciplines which require similar skillsets, but have more representative student populations, to participate in the Government-backed PhD programme.**
175. **We acknowledge the considerable scepticism of at least some technology companies who believe that the apprenticeship levy is of little use to them, despite the success that others in the sector have had with apprenticeships. The Government should produce clear guidance on how the apprenticeship levy can be best deployed for use in the technology sector, in particular in SMEs and start-ups.**

Immigration and overseas skilled workers

176. While a majority of witnesses believed measures, such as the funding of additional PhD places, should be taken to develop the UK's home-grown AI talent, this was generally seen as a long-term solution, which would not address shortages for some years to come. Many witnesses highlighted the importance of overseas workers to the UK AI development sector, and voiced concerns that this supply could be jeopardised by illiberal immigration policies, especially in the wake of Brexit. The think tank Future Advocacy observed that:

“In the current climate of uncertainty, there has already been a sharp decline in EU applications to UK tech jobs. 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’”.²²⁸

177. Those with machine learning expertise are globally sought after, and therefore constitute a highly mobile population. Eileen Burbidge emphasised that the “uncertainty (of Brexit) alone raises issues and gives people pause before they consider coming to the UK, or challenges for companies trying to recruit outside the UK”. Professor Wooldridge warned that Brexit “could quite genuinely be the death knell for UK tech start-ups, which are heavily reliant on overseas talent”.²²⁹
178. A 2017 assessment by Tech City found that non-UK workers made up 13% of the digital technology workforce, compared to 10% in the wider economy in 2015. Interestingly, non-EU workers accounted for a larger share of employment (7%) in the digital technology industries than EU workers (6%)—however, employment for EU nationals had grown faster than for non-EU nationals, growing by two percentage points over the five years from 2011 to 2015.²³⁰

228 Written evidence from Future Advocacy (AIC0121)

229 Q 47 (Eileen Burbidge) and written evidence from Professor Michael Wooldridge (AIC0174)

230 Tech City, ‘The nationality of workers in the UK tech industry – Tech Nation Talent: Part 1’: <http://www.techcityuk.com/blog/2017/10/tech-nation-talent-nationality-workers-uk-tech-industry/> [accessed 31 January 2018]

179. Many witnesses called on the Government to confirm the position of EU technology workers after Brexit, and to liberalise the visa regimes for overseas technology workers in the UK more generally. Balderton Capital suggested that special measures should be taken: “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”.²³¹ Eileen Burbidge also asked that the Home Office Migration Advisory Committee consider adding artificial intelligence-related roles to the Tier 2 Shortage Occupation List, and that the quota on Tier 1 (Exceptional Talent) visas be increased.²³²
180. On 15 November, shortly after we spoke with Eileen Burbidge, the Government did indeed announce that it would be doubling the number of Tier 1 (Exceptional Talent) visas from 1,000 to 2,000 a year. The 2,000 visas will be made available to individuals who are “recognised as existing global leaders or promising future leaders in the digital technology, science, arts and creative sectors” by one of five UK endorsing organisations:
- Tech City UK
 - Arts Council England
 - The British Academy
 - The Royal Society
 - The Royal Academy of Engineering.²³³
181. **The Government’s announcement that it will increase the annual number of Tier 1 (exceptional talent) visas from 1,000 to 2,000 per year is welcome. While top-tier PhD researchers and designers are required, a thriving AI development sector is also dependent on access to those able to implement artificial intelligence research, whose occupations may fall short of the exceptional talent requirements.**
182. **We are concerned that the number of workers provided for under the Tier 1 (exceptional talent) visa scheme will be insufficient and the requirements too high level for the needs of UK companies and start-ups. We recommend that the number of visas available for AI researchers and developers be increased by, for example, adding machine learning and associated skills to the Tier 2 Shortage Occupations List.**

231 Written evidence from Balderton Capital (UK) LLP ([AIC0232](#))

232 [Q 54](#) (Eileen Burbidge). In the UK immigration system, Tier 1 (Exceptional Talent) visas are for people recognised by the Home Office as a leader (or emerging leader) in their field. Tier 2 visas allow skilled workers to enter the UK on a long term basis to fill a skilled job vacancy (defined by an occupation list). Tier 2 jobs must usually be advertised to workers from within the European Economic Area (EEA) before they can be offered to those from outside the EEA. The Tier 2 Shortage Occupation List is a list of occupations which UK employers have not been able to recruit, and the jobs on this list can be offered to people from beyond the EEA, without being advertised within the EEA first.

233 Home Office, ‘Government doubles exceptional talent visa offer’ (15 November 2017): <https://www.gov.uk/government/news/government-doubles-exceptional-talent-visa-offer> [accessed 7 February 2018]

Maintaining innovation

183. While deep learning has played a large part in the impressive progress made by AI over the past decade, it is not without its issues. Some of our witnesses believed it would not continue to deliver advances at the current rate, and that other avenues of research needed more support. Deep learning requires large datasets, which can be difficult and expensive to obtain, and can require a great deal of processing power. As a number of witnesses emphasised, recent advances in deep learning have been made possible with the growth of cheap processing power.²³⁴ Some indications suggest, however, that Moore’s law—the observation that the number of transistors on a circuit board tends to double every two years—is starting to break down, with the growth in cheap processing power slowing.²³⁵ Innovations such as quantum computing may yet restore, or even accelerate, the historic growth in cheap processing power, but it is currently too early to say this with any certainty.²³⁶
184. Several of our witnesses also pointed to the difficulties of transfer learning, or “the ability of computers to infer what might work in a given scenario based on knowledge gained in an apparently unrelated scenario”.²³⁷ This can also be thought of as common sense in human beings, which cannot currently be replicated in any AI system.²³⁸
185. The Foundation for Responsible Robotics believed it could be the case that “once all the low hanging fruit has been picked, severe limitations will be found and then the technology will plateau”.²³⁹ Geoff Hinton, a pioneer in deep learning who is still revered in the field today, has warned that the deep learning revolution might be drawing to a close.²⁴⁰ Others were more optimistic. Witnesses told us of advances in custom-designed ‘AI chips’,²⁴¹ such as Google’s Tensor Processing Unit, which trades general-purpose processing power for extra power in AI applications and, more speculatively, advances in quantum computing which might provide further boosts in future.²⁴² In terms of transfer learning, DeepMind has had success in applying its AlphaGo AI, originally trained on Go, to chess, but for the most part this is still challenging.
186. Some of the most innovative AI research we observed went beyond deep learning, and combined it with other aspects of AI, creating hybrid systems which sought to compensate for the drawbacks of any one style of AI. For example, Prowler.IO, an AI start-up in Cambridge, told us of their own issues with deep learning, in particular the amount of data it often requires,

234 Written evidence from Capco (AIC0071); Dr Toby Walsh (AIC0078); Economic Singularity Supper Club (AIC0058) and BioCentre (AIC0169)

235 Tony Simonite, ‘How AI Can Keep Accelerating After Moore’s Law’, *MIT Technology Review* (30 May 2017): <https://www.technologyreview.com/s/607917/how-ai-can-keep-accelerating-after-moores-law/> [accessed 8 February 2018]; Mark Pesce, ‘Death notice: Moore’s Law. 19 April 1965–2 January 2018’, *The Register* (24 January 2018): https://www.theregister.co.uk/2018/01/24/death_notice_for_moores_law/ [accessed 8 February 2018]

236 George Musser, ‘Job one for quantum computers: boost artificial intelligence’, *Wired* (10 February 2018): <https://www.wired.com/story/job-one-for-quantum-computers-boost-artificial-intelligence/> [accessed 14 February 2018]

237 Written evidence from Royal Academy of Engineering (AIC0140) See also Touch Surgery (AIC0070)

238 Written evidence IBM (AIC0160); Q 2 (Professor Nick Bostrom) and The Reverend Dr Lyndon Drake (AIC0108)

239 Written evidence from Foundation for Responsible Robotics (AIC0188)

240 James Somers, ‘Is AI riding a one-trick pony?’, *MIT Technology Review* (29 September 2017): <https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony/> [accessed 8 February 2018]

241 Written evidence from Research Council UK (AIC0142) and Deep Learning Partnership (AIC0027)

242 Written evidence from BioCentre (AIC0169) and Economic Singularity Supper Club (AIC0058)

and the lack of transparency behind the decisions it comes to. They outlined their own approach to us, which combines a range of approaches, including probabilistic modelling, multi-agent systems and reinforcement learning to create more robust AI which can cope with less data and more uncertainty. A number of other AI experts have also suggested that combining different approaches, some of which were once prominent within the AI field but have since fallen out of fashion relative to deep learning, may well be a productive way forward.²⁴³

187. Google told us that the success of AI in the UK was in considerable part due to the Government's traditional role in "supporting long term fundamental research", and that this role should continue.²⁴⁴ Professor Wolfgang Wahlster also emphasised the UK's role in pioneering AI research.²⁴⁵ It will also be important for the UK to continue to participate in European research and innovation programmes such as Horizon 2020, and its successor Framework 9. We welcome the Government's commitment to underwrite any bids for Horizon 2020 projects while the UK is still a member of the EU, and we hope that this continues, where possible, after we leave the European Union.²⁴⁶
188. **We believe that the Government must commit to underwriting, and where necessary replacing, funding for European research and innovation programmes, after we have left the European Union.**
189. We should also consider this approach to research in the light of the evidence we received on deep learning in relation to other aspects of AI. As Jonathan Penn, a historian of AI at the University of Cambridge, told us, AI has long been a varied and even incoherent field at times, with different sub-disciplines constantly vying for attention and funding. Neural networks were, for example deemed a "sterile" area of research by Marvin Minsky, a view shared by many in the discipline for many decades.²⁴⁷ Geoff Hinton renewed interest in neural networks just as most AI researchers were fully investing in the now largely defunct area of expert systems.
190. Furthermore, in order to track and assess the likely impact of AI on the economy, work, politics, health care and medicine, education and other fields, it will be crucial for experts from different disciplines to work closely together. In particular, researchers specialising in AI will need to collaborate with those studying other academic areas. Institutes such as the Leverhulme Centre for the Future of Intelligence at the University of Cambridge, and the Oxford Internet Institute at the University of Oxford are excellent, existing, examples of this collaboration, and universities across the UK should encourage the development of their own such centres.

243 Richard Waters, 'Why we are in danger of overestimating AI', *Financial Times* (5 February 2018): <https://www.ft.com/content/4367e34e-db72-11e7-9504-59efdb70e12f> [accessed 6 February 2018]

244 Written evidence from Google (AIC0225)

245 Q 168 (Professor Wolfgang Wahlster)

246 Department for Exiting the European Union, *Collaboration on science and innovation: a future partnership paper* (6 September 2017): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/642542/Science_and_innovation_paper.pdf [accessed 1 March 2018]

247 Written evidence from Jonathan Penn (AIC0198)

191. **The state has an important role in supporting AI research through the research councils and other mechanisms, and should be mindful to ensure that the UK's advantages in AI R&D are maintained. There is a risk that the current focus on deep learning is distracting attention away from other aspects of AI research, which could contribute to the next big advances in the field. The Government and universities have an important role to play in supporting diverse sub-fields of AI research, beyond the now well-funded area of deep learning, in order to ensure that the UK remains at the cutting edge of AI developments.**

CHAPTER 5: WORKING WITH ARTIFICIAL INTELLIGENCE

192. The economic impact of AI in the UK could be profound. This chapter considers two widely shared concerns for policymakers, businesses and the general public: the UK's productivity puzzle, and the potential impact of AI on the labour market.

Productivity

193. The opportunity that the widespread use of artificial intelligence offers to improve productivity in the UK was, perhaps, the most common benefit cited to us by our witnesses. Andrew de Rozairo, Vice President of Customer Innovation and Enterprise Platform, SAP, said “if we adopt AI, given the strong skillsets that we have in the UK, we have a huge opportunity to boost productivity”.²⁴⁸ TechUK told us: “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”.²⁴⁹ The Government also recognised this benefit: “Impacts in industry ... are likely to be profound in terms of productivity”.²⁵⁰ A number of witnesses argued that productivity would be improved as human labour was augmented in a range of ways, such as summarising complex documents or sorting email inboxes.²⁵¹
194. However, a note of caution is advisable, as the economic consequences of earlier phases of computerisation, particularly in the 1970s and 1980s, are still poorly understood. Some economists have argued that there is still little evidence that information technology had a significant impact on productivity over this period, especially in the United States, a phenomenon which famously led economist Robert Solow to remark that “you can see the computer age everywhere but in the productivity statistics”.²⁵² Others argue that increased productivity in the 1990s show that gains were merely delayed.²⁵³ Either way, prior experiences with computerisation suggest that any relationship between AI adoption and productivity is unlikely to be necessary or straightforward in nature.

Box 6: What is productivity?

Productivity measures how efficiently work is converted into the output of goods and services. The better the productivity, the more goods and services are being produced per hour worked. Productivity is used to assess economic growth and competitiveness, and as the basis for international comparison of a country's performance. In the UK, the Office for National Statistics regularly reports on labour productivity, based on output per hour, output per job, and output per worker.

248 Q 84 (Andrew de Rozairo)

249 Written evidence from techUK (AIC0203)

250 Written evidence from Department for Digital, Culture, Media and Sport and Department for Business Energy and Industrial Strategy (AIC0229)

251 Written evidence from Fujitsu (AIC0120); Information Technology Industry Council (AIC0176); Imperial College London (AIC0214) and Arm (AIC0083)

252 Daron Acemoglu, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price, ‘Return of the Solow Paradox? IT, Productivity, and Employment in US Manufacturing’ in *American Economic Review*, vol. 104 (2014), pp 394–399: <http://economics.mit.edu/files/10414> [accessed 14 February 2014]

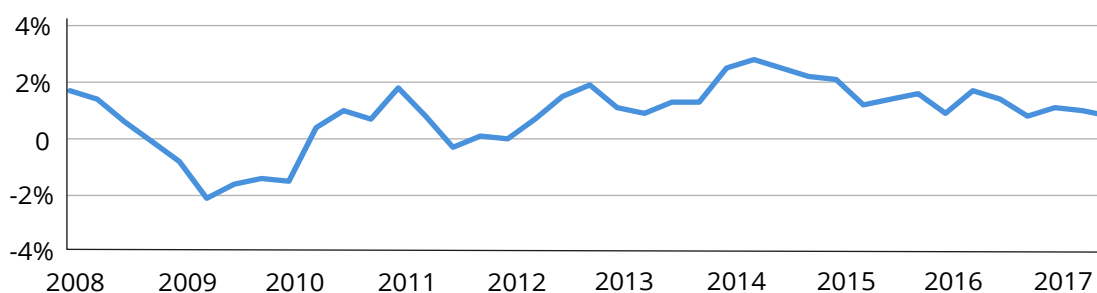
253 Erik Brynjolfsson, Shinkyu Yang, ‘The Intangible Costs and Benefits of Computer Investments: Evidence from the Financial Markets’, *MIT Sloan School of Management* (December 1999): <http://digital.mit.edu/erik/itq00-11-25.pdf> [accessed 14 February 2018]

195. Hopes that AI will improve productivity must be set against a backdrop of low productivity growth across the developed world, and almost non-existent productivity growth in the UK since the 2008 financial crisis. The Royal Society for the encouragement of Arts, Manufacturers and Commerce (RSA) told us 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”.²⁵⁴ Sarah O’Connor said this was “a puzzle that is taxing the best minds in technology and economics right now”.²⁵⁵ She also told us that “if you do not have productivity growing at a decent clip then you cannot have sustainable increases in living standards” and AI “could mean a step change in productivity”.²⁵⁶ The Center for Data Innovation put it in even more stark terms, saying that finding a way to improve productivity was:

“ ... 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”.²⁵⁷

196. The Office for National Statistics reported in January 2018 that productivity had grown by 0.9% from Quarter 2 (April to June) 2017 to Quarter 3 (July to September) 2017—this is the largest increase in productivity since Quarter 2 2011. This is, however, not a significant enough improvement to overlook the potential benefits that AI might have for productivity in the UK.

Figure 4: UK job productivity



Source: Office for National Statistics, ‘Productivity jobs: whole economy: % changer per annum:SA: UK’ (5 January 2018): <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/labourproductivity/timeseries/lnno/prdy> [accessed 7 February 2017]

197. We share the optimism of our witnesses that AI could improve productivity. We also share their concerns that this is an opportunity that could be missed in the UK. The Institute of Chartered Accountants in England and Wales (ICAEW) said: “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”.²⁵⁸ Sage told us that their research demonstrated “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 and medium-sized business”.²⁵⁹ The research suggested that the amount of time spent on such tasks is because digital tools have not been

254 Written evidence from The RSA ([AIC0157](#))

255 [Q 10](#) (Sarah O’Connor)

256 *Ibid.*

257 Written evidence from Center for Data Innovation ([AIC0043](#))

258 Written evidence from the Institute of Chartered Accountants in England and Wales (ICAEW) ([AIC0041](#))

259 Written evidence from Sage ([AIC0159](#))

adopted, and that if UK businesses could be 5% more productive, GDP could increase by £33.9 billion per year.²⁶⁰ Kriti Sharma told us that “lack of digital adoption is leading to what we call the productivity gap in the UK”.²⁶¹

198. In 2017, 5.7 million businesses in the UK were classified as SMEs (99% of all businesses), with 5.4 million of those employing fewer than 10 staff. Such micro-businesses accounted for 33% of employment and 22% of turnover.²⁶² The CBI, which represents over 190,000 businesses in the UK, said: “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”.²⁶³ If businesses are not using existing technology, in particular SMEs, we are concerned that the potential benefits to productivity offered by artificial intelligence could bypass significant portions of the business community in the UK.
199. **We support the Government’s belief that artificial intelligence offers an opportunity to improve productivity. However, to meet this potential for the UK as a whole, the AI Council must take a role in enabling AI to benefit all companies (big and small) and ensuring they are able to take advantage of existing technology, in order for them to take advantage of future technology. It will be important that the Council identifies accelerators and obstacles to the use of AI to improve productivity, and advises the Government on the appropriate course of action to take.**
200. Other witnesses shared concerns about the state of digital infrastructure in the UK. Research Councils UK said:
- “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”.²⁶⁴
201. Vishal Wilde said “those who live in rural areas and who do not have access to broadband also do not feel the benefits of ... the productivity gains associated with AI nearly as much as other parts of the country”.²⁶⁵ Artificial intelligence is, in part, reliant on access to digital infrastructure. Without the digital foundations (both physical in terms of internet connectivity and in terms of skills, discussed later in this report) the potential benefits of artificial intelligence to the UK’s productivity will be neither realised nor widespread.

Box 7: Broadband speeds

Superfast broadband is defined, by Ofcom, as connections providing download speeds in excess of 30 megabits per second (Mbps).

Ultrafast broadband is considered to be where speeds are in excess of 300 Mbps.

Source: Ofcom, *Connected Nations 2017* (15 December 2017): https://www.ofcom.org.uk/_data/assets/pdf_file/0024/108843/summary-report-connected-nations-2017.pdf [accessed 14 February 2018]

260 *Ibid.*

261 Q 77 (Kriti Sharma)

262 House of Commons Library, Business Statistics, Briefing Paper, [Number 06152](#), December 2017

263 Written evidence from CBI ([AIC0114](#))

264 Written evidence from Research Councils UK ([AIC0142](#))

265 Written evidence from Vishal Wilde ([AIC0004](#))

202. To improve national infrastructure, the Industrial Strategy sets out plans to increase the National Productivity Investment Fund from £23 billion to £31 billion, and to improve digital infrastructure with over £1 billion of public investment, including £176 million for 5G and £200 million for full-fibre broadband networks.²⁶⁶ Matt Hancock MP told us that 95% of premises would have access to superfast broadband by the end of 2017, and that ultrafast connectivity rollout is the next ambition for the Government. This would be delivered via “a competitive market with many players bringing ultrafast speeds over full-fibre technology”.²⁶⁷ As of February 2018, 3% of the UK was covered by full-fibre broadband.²⁶⁸
203. **We welcome the Government’s intentions to upgrade the nation’s digital infrastructure, as far as they go. However, we are concerned that it does not have enough impetus behind it to ensure that the digital foundations of the country are in place in time to take advantage of the potential artificial intelligence offers. We urge the Government to consider further substantial public investment to ensure that everywhere in the UK is included within the rollout of 5G and ultrafast broadband, as this should be seen as a necessity.**

Government adoption, and procurement, of artificial intelligence

204. Our witnesses reminded us of the importance of government as a customer, nationally and locally. It can both procure AI solutions for the public sector and adopt the technology, thereby supporting UK-based technology companies.
205. BSA (The Software Alliance), a global software advocate, said: “The UK Government could help demonstrate AI’s potential benefits by investing in innovative AI implementations in the public sector”.²⁶⁹ Professor Susskind told us that “in the public sector ... use of AI and other advanced technologies should transform and not simply streamline our current ways of working and governing”.²⁷⁰ Doteveryone said: “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”.²⁷¹ TechUK suggested that “the use of AI virtual agents across Government departments and the public sector could save an estimated £4 billion a year”.²⁷² Microsoft argued that deployment of artificial intelligence in the public sector could enable more informed policy decisions, and innovative uses of AI could help address public and societal challenges.²⁷³
206. The Hall-Pesenti Review recommended that:

“Government, drawing on the expertise of the Government Digital Service, the Data Science Partnership and experts working with data in other Departments, should develop a programme of actions to prepare

266 *Industrial Strategy: Building a Britain fit for the future*, p 39

267 [Q 190](#) (Matt Hancock MP)

268 Ofcom, ‘New Ofcom rules to boost full-fibre broadband’ (23 February 2018): <https://www.ofcom.org.uk/about-ofcom/latest/features-and-news/new-rules-boost-full-fibre-broadband> [accessed 5 March 2018]

269 Written evidence from BSA The Software Alliance ([AIC0153](#))

270 Written evidence from Professor Richard Susskind ([AIC0194](#))

271 Written evidence from Doteveryone ([AIC0148](#))

272 Written evidence from techUK ([AIC0203](#))

273 Written evidence from Microsoft ([AIC0149](#))

the public sector and spread best practice for applying AI to improve operations and services for citizens”.²⁷⁴

207. The Government shared with us as part of their written evidence the AI tools and programmes it is using, or looking to use, in the near future.²⁷⁵ We welcome the fact that Government departments are actively considering the use of artificial intelligence in the delivery of public services, in particular using innovative approaches such as *Kaggle* competitions.²⁷⁶ Other governments are already doing this. For example, in December 2017, the US Department of Homeland Security offered \$1.5 million in prizes for their ‘Passenger Screening Algorithm Challenge’, which aimed to improve the accuracy of threat prediction algorithms used in airport security. Datasets were made available and the challenge ran for about a fortnight. Such innovative crowdsourcing can help governments and businesses tap into world-class expertise which they would otherwise not be able to access.
208. However, the Government could still do more to deploy AI and we are conscious that it is considering this. The Autumn Budget in 2017 announced the establishment of the GovTech Catalyst. This will be a small unit based within the Government Digital Service which will provide a direct access point to Government for businesses and innovators. The GovTech fund is £20 million over three years to support public bodies in procuring innovative products. Matt Hancock MP told us “most of GovTech is about procurement ... Indeed, getting procurement rules right is one of the most important parts of driving improvements in technology through government, because you need the leadership and the permission from the top to drive the change”.²⁷⁷
209. The Royal Society agreed with the Minister’s assessment:
- “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”.²⁷⁸
210. The Government spends £45 billion a year on procuring goods and services.²⁷⁹ As such, it is one of the most significant ‘customers’ in the United Kingdom, and has immense power in encouraging the adoption of new behaviours and practices in its supply chains.

274 Department for Business, Energy & Industrial Strategy and Department for Digital, Culture, Media and Sport, ‘Recommendations of the review’ (15 October 2017): <https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk/recommendations-of-the-review> [accessed 1 February 2018]

275 Written evidence from HM Government (AIC0229)

276 *Kaggle* is an online platform, owned by Google, which hosts data science and machine learning competitions to which data and computer scientists compete to develop the best models for handling provided datasets. Rewards are offered (often in terms of cash prizes) for these solutions.

277 Q 192 (Matt Hancock MP)

278 Written evidence from Royal Society (AIC0168)

279 National Audit Office, *Government’s spending with small and medium-sized enterprises* (March 2016): <https://www.nao.org.uk/wp-content/uploads/2016/03/Governments-spending-with-small-and-medium-sizes-enterprises.pdf> [accessed 12 January 2018]

211. In the UK, the Crown Commercial Service, an agency of the Cabinet Office, is responsible for procurement policy for the Government, enabling cost-efficient procurement by bringing together requests for the same goods or services, as well as supporting smaller projects. The Crown Commercial Service works with over 17,000 customer organisations in the public sector and has more than 5,000 suppliers, and thereby has significant influence over various sectors in the country.²⁸⁰
212. Our witnesses suggested that Government procurement could be used to encourage greater adoption of artificial intelligence, both through the companies contracting directly with Departments and via the Crown Commercial Service. SCAMPI, a research project at City, University of London, said: “the UK public sector ... is currently benefiting little from the development and use of artificial intelligence, as few initiatives have been funded or reported”.²⁸¹ The UK Computing Research Committee said “the public sector could do more to benefit from these techniques to support the provision and optimisation of services across a host of areas”.²⁸²
213. Public procurement in the UK is subject to the Treaty on the Function of the European Union’s (TFEU) principles of non-discrimination, the free movement of goods, the freedom to provide services and the freedom of establishment. This is realised via a series of directives,²⁸³ which have been translated into domestic law. In the UK, this means that central Government, and other public organisations, must advertise contracts for goods, services and works which are worth over £10,000 at a UK-wide level, and at an EU-wide level for services and supplies contracts worth over £118,000. The threshold for advertising ‘works’ contracts across the EU is £4.5 million.²⁸⁴ Given the UK’s departure from the European Union, there is an opportunity for the Crown Commercial Service to ensure that these rules and thresholds benefit businesses in the UK, in particular when it comes to public sector procurement and the stimulation of a fertile AI development sector, as long as it is still a competitive process.
214. In 2013, the Government Digital Service launched the Service Design Manual, intended to help improve public services and ensure the adoption of digital approaches wherever possible. The Manual includes guidance and instructions on how to approach choosing technology, and use the Technology Code of Practice²⁸⁵ as part of the spend control process. The Technology Code of Practice includes points such as “use cloud first” and “make better use of data”. There is no explicit mention of artificial intelligence in the Code. For the Government to adopt artificial intelligence in the delivery of public services, the need to consider AI solutions must be embedded in the decision-making process right from the start. As such, amending the Code could be one approach to ensure that technologists in the civil service actively consider the use of artificial intelligence.

280 Crown Commercial Service, ‘About us’: <https://www.gov.uk/government/organisations/crown-commercial-service/about> [accessed 14 February 2018]

281 Written evidence from SCAMPI Research Consortium, City, University of London (AIC0060)

282 Written evidence from UK Computing Research Committee (AIC0030)

283 See Public Sector: Directive 2014/24/EU, Concessions: Directive 2014/23/EU and Utilities: Directive 2014/25/EU

284 Crown Commercial Service, *Procurement Policy Note—New Thresholds 2018* (December 2017): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/670666/PPN_0417_New_Thresholds_2018_1.pdf [accessed 17 January 2018]

285 The Government’s criteria for designing, building and buying better technology.

215. **The Government’s leadership in the development and deployment of artificial intelligence must be accompanied by action. We welcome the announcement of the GovTech Catalyst and hope that it can open the doors of Whitehall to the burgeoning AI development sector in the UK. We also endorse the recommendation of the Hall-Pesenti Review aimed at encouraging greater use of AI in the public sector.**
216. **To ensure greater uptake of AI in the public sector, and to lever the Government’s position as a customer in the UK, we recommend that public procurement regulations are reviewed and amended to ensure that UK-based companies offering AI solutions are invited to tender and given the greatest opportunity to participate. The Crown Commercial Service, in conjunction with the Government Digital Office, should review the Government Service Design Manual and the Technology Code of Practice to ensure that the procurement of AI-powered systems designed by UK companies is encouraged and incentivised, and done in an ethical manner.**
217. **We also encourage the Government to be bold in its approach to the procurement of artificial intelligence systems, and to encourage the development of possible solutions to public policy challenges through limited speculative investment and support to businesses which helps them convert ideas to prototypes, in order to determine whether their solutions are viable. The value of AI systems which are deployed to the taxpayer will compensate for any money lost in supporting the development of other tools.**
218. **Finally, with respect to public procurement, we recommend the establishment of an online bulletin board for the advertisement of challenges which the Government Office for AI and the GovTech Catalyst have identified from across Government and the wider public sector where there could be the potential for innovative tech- and AI-based solutions.**

Impact on the labour market

219. The potential impact of AI on the wider economy was one of the most widely discussed and contentious issues of our inquiry. The prospect of significant productivity gains from AI invariably raises the prospect of increased unemployment, although it is equally possible that productivity can grow alongside employment, assuming economic output also increases. The proportion of jobs estimated to be at risk in developed economies such as the UK normally range between 10% and 50%, over the next 10–20 years. The kinds of jobs most at risk are also frequently debated, with some arguing that low-skilled jobs are at far greater risk, while others argue that many white-collar, but relatively repetitive or less creative jobs, might also be at risk.
220. As a number of our witnesses have emphasised, while the current debate stems largely from academic work carried out at the start of this decade, concerns surrounding ‘technological unemployment’ have a long and distinguished history. The 1920s and 1930s saw much public debate on both sides of the Atlantic about the threat of ‘technological unemployment’, a term popularised by John Maynard Keynes. In 1949 Norbert Wiener warned that computerisation, combined with “the valuation of human beings on which our present factory system is based” could usher in “an industrial

revolution of unmitigated cruelty”.²⁸⁶ In the 1960s similar anxieties led to President Johnson establishing the US National Commission on Technology, Automation and Economic Progress in 1964. Professor Edgerton noted the similarities here in the UK, when in 1963 Harold Wilson warned that “computers have reached the point where they command facilities of memory and of judgment far beyond the capacity of any human being or group of human beings who have ever lived”, and speculated that the white collar professions would be particularly hard hit.²⁸⁷

221. Contemporary concerns can largely be traced back to Erik Brynjolfsson and Andrew McAfee’s influential 2011 book, *Race Against the Machine*, which predicted widespread disruption and upheaval as a result of accelerating automation, in part as a consequence of advances in AI.²⁸⁸ In 2013 Carl Frey and Michael Osborne started a trend for attempting more precise predictions, and by examining the jobs they believed were most susceptible to automation with current or near-future technology, they claimed that around 47% of total US employment was at risk of automation.²⁸⁹ Jeremy Bowles, applying the same methodology, calculated that 54% of jobs across the EU were similarly threatened.²⁹⁰
222. In a 2016 study, Organisation for Economic Co-operation and Development (OECD) economists devised a further methodological innovation, and shifted their attention to focus on how automatable particular tasks within jobs were. While some current jobs may be composed solely of tasks which are completely automatable, therefore rendering the job itself automatable, they concluded that most jobs did not currently fall into this category. Using this methodology, John Hawksworth and Richard Berriman concluded in a 2017 report for PwC that up to 30% of existing UK jobs are at ‘high risk’ of automation by the 2030s.²⁹¹ These risks are highest in sectors such as transportation and storage (56%), manufacturing (46%) and wholesale and retail (44%). However, Berriman and Hawksworth believe that due to the additional jobs which are likely to be created through economic growth over this period, the net effect on employment is likely to be neutral.
223. Over the same period, some economists have adopted a more historical approach, by focusing on patterns of automation in the recent past. These include Daron Acemoglu and Pascual Restrepo’s study of the impact of industrial robot usage between 1990 and 2007 on the US labour market, and David Autor’s work on the history and future of workplace automation. While Acemoglu and Restrepo have estimated that areas in the US most exposed to industrial automation in the 1990s and 2000s experienced “large and

286 John Markoff, ‘In 1949, He Imagined an Age of Robots’, *The New York Times* (20 May 2013): http://www.nytimes.com/2013/05/21/science/mit-scholars-1949-essay-on-machine-age-is-found.html?pagewanted=all&_r=0 [accessed 30 January 2018]

287 Q 214 (Professor David Edgerton)

288 Andrew McAfee and Erik Brynjolfsson, *Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy* (Lexington, Massachusetts: Digital Frontier Press, 2011)

289 Carl Frey and Michael Osborne, *The future of employment: How susceptible are jobs to computerisation?* (17 September 2013): https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf [accessed 1 February 2018]

290 Jeremy Bowles, ‘The computerisation of European jobs’, *bruegel.org* (24 July 2017): <http://bruegel.org/2014/07/the-computerisation-of-european-jobs/> [accessed 1 February 2018]

291 Richard Berriman, John Hawksworth, ‘Will robots steal our jobs? The potential impact of automation on the UK and other major economies’, *PwC UK Economic Outlook* (March 2017): <https://www.pwc.co.uk/economic-services/ukeyo/pwcukeyo-section-4-automation-march-2017-v2.pdf> [accessed 1 February 2018]

robust negative effects” on employment and wages, Autor concludes quite the opposite, noting the ways in which automation has often complemented human labour in ways which “increase productivity, raise earnings and augment the demand for labour”.²⁹²

224. This research has in turn percolated through the policymaking world, influencing a number of important recent reports on the impact of AI and automation on the labour market. The RSA have emphasised the need to consider the impact of AI and automation on the quality and substance of jobs, particularly low-skilled jobs, and the need “to accelerate the adoption of AI and robotics ... in a way that delivers automation on our own terms”.²⁹³ Future Advocacy, a London-based think tank, have argued that while the net impact of AI and automation might be relatively neutral, the impact across different regions of the UK is likely to be highly divergent based on their current economic strengths and weaknesses.²⁹⁴ The IPPR’s report in December 2017 on the subject, drawing heavily on Autor’s approach, argued that automation is likely to transform, rather than eliminate, work, but that policies will be needed to both to accelerate automation in the interests of boosting productivity and wages, and manage the growing inequalities in wealth, income and power which could otherwise arise.²⁹⁵
225. Evidence we received varied on the likely nature and severity of this impact. One school of thought, generally favoured by businesses and those developing AI systems, believed that the impact would be relatively moderate, or even positive.²⁹⁶ Tasks, rather than entire jobs, were likely to be automated, and therefore human capacities in many jobs would be augmented, rather than replaced.²⁹⁷ As the Canadian Institute for Advanced Research’s (CIFAR) evidence explained, “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)”.²⁹⁸ Others argued that even if many types of jobs were entirely automated (AI as a ‘replacing technology’), other jobs would be created in the process, as happened during the nineteenth-century industrial revolution.²⁹⁹ Very few witnesses provided much detail on what these new jobs might look like, although they may well be impossible to predict.

292 Daron Acemoglu and Pascual Restrepo, ‘Robots and Jobs: Evidence from US Labor Markets’ *NBER Working Paper No w23285* (March 2017): <https://economics.mit.edu/files/12763>; David Autor, ‘Why are there still so many jobs? The history and future of workplace automation’ in *The Journal of Economic Perspectives*, vol. 29, no. 3 (2015), p 5: <https://economics.mit.edu/files/11563> [accessed 8 March 2019]

293 The RSA, *The Age of Automation* (September 2017), p 8: https://www.thersa.org/globalassets/pdfs/reports/rsa_the-age-of-automation-report.pdf [accessed 17 January 2018]

294 Future Advocacy, *The impact of AI in UK constituencies: Where will automation hit hardest?* (October 2017): <https://static1.squarespace.com/static/5621e990e4b07de840c6ea69/t/59e777fcd7bdce3041b57ac3/1508341775530/FutureAdvocacy-GeographicalAI.pdf> [accessed 1 February 2018]

295 IPPR, *Managing automation: Employment, inequality and ethics in the digital age* (28 December 2017), p 20: <https://www.ippr.org/files/2017-12/cej-managing-automation-december2017-1-.pdf> [accessed 17 January 2018]

296 See, for example, written evidence from the RSA (AIC0157); Dr Ian Morgan and Brian Joyce (AIC0179) and euRobotics Topics Group on ‘Ethical, Legal and Socio-economic issues’ (AIC0189)

297 See, for example, written evidence from Fujitsu (AIC0120); Information Technology Industry Council ITI (AIC0176); Imperial College London (AIC0214) and Arm (AIC0083)

298 Written evidence from CIFAR (AIC0136)

299 See, for example, written evidence from Braintree (AIC0074); CBI (AIC0114); CIFAR (AIC0136); Fujitsu (AIC0120); Innovate UK (AIC0220); Microsoft (AIC0149); BSA The Software Alliance (AIC0153); The RSA (AIC0157) and euRobotics Topics Group on ‘Ethical, Legal and Socio-economic issues’ (AIC0189)

226. The other broad school of thought came from think tanks and NGOs, and proposed that AI was likely to be far more disruptive to future employment patterns, as many blue- and white-collar jobs might be automated over a very short space of time, hindering the chances for those made redundant to find alternative work.³⁰⁰ Such witnesses warned that the impact of such change would not be evenly distributed across the country. One witness, with experience in the call centre industry, emphasised the potential scale of the challenge in their own industry:

“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 8.15 billion calls to UK Contact Centres of which there are 7,500 employing just under one 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”.³⁰¹

227. However, considerable doubts were raised by later witnesses in our inquiry regarding the methodological soundness of much of the academic literature in this area. Professor Susskind, noting that “there is no evidence from the future”, emphasised that studies which have broken down jobs into their constituent tasks could be misleading, as jobs have frequently been reconstituted around new technologies before, and automated processes do not generally entail simply copying jobs or professions as they existed prior to automation.³⁰² As a result, Professor Susskind found many of the predictions about job losses to be “entirely unreliable”, and predicting job creation “even harder”, and dismissed the “quasi-science of the major consulting firms” as lacking deep theoretical foundations. Professor Dame Henrietta Moore, Director of the Institute for Global Prosperity, UCL, and Olly Buston, CEO and Founder, Future Advocacy, while still believing in the need for action to counter the possibility of job losses, concurred that many predictions were “evidence light”.³⁰³

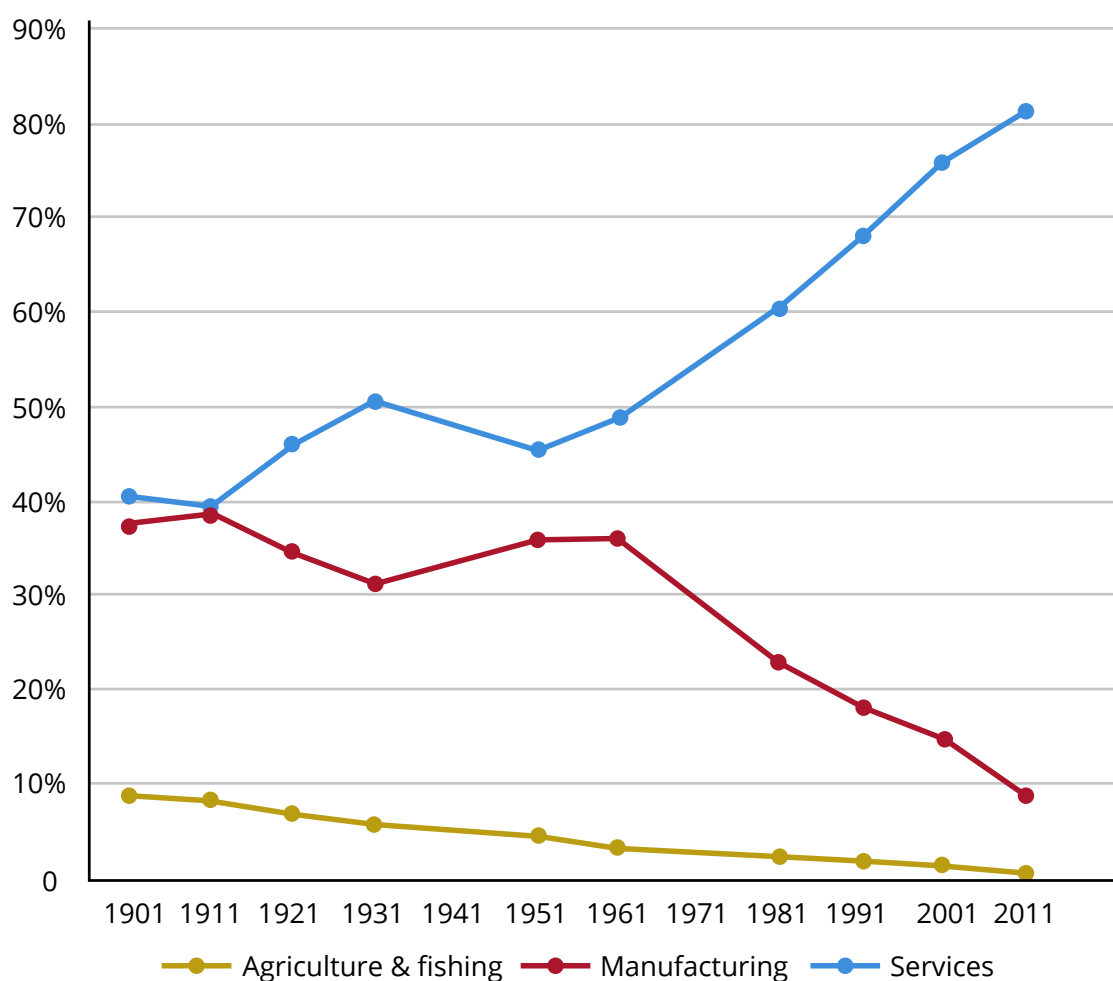
300 See, for example, written evidence from Charities Aid Foundation ([AIC0042](#)); The Economic Singularity Supper Club ([AIC0058](#)); The Knowledge, Skills and Experience Foundation ([AIC0044](#)); Professor Maja Pantic ([AIC0215](#)); Dr Mike Lynch ([AIC0005](#)); Warwick Business School, University of Warwick ([AIC0117](#)); Deep Science Ventures ([AIC0167](#)) and Dr Aysegul Bugra, Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis and Dr Valentina Rita Scotti ([AIC0051](#))

301 Written evidence from Contact Centre Systems Ltd. ([AIC0032](#))

302 [Q 97](#) (Professor Richard Susskind)

303 [Q 97](#) (Professor Henrietta Moore and Olly Buston)

Figure 5: Percentage of working people employed in each industry group, 1901–2011



Source: ONS, '2011 census analysis, 170 years of industry' (2013): <http://webarchive.nationalarchives.gov.uk/20160108022535/http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-309799> [accessed 28 February 2018]³⁰⁴

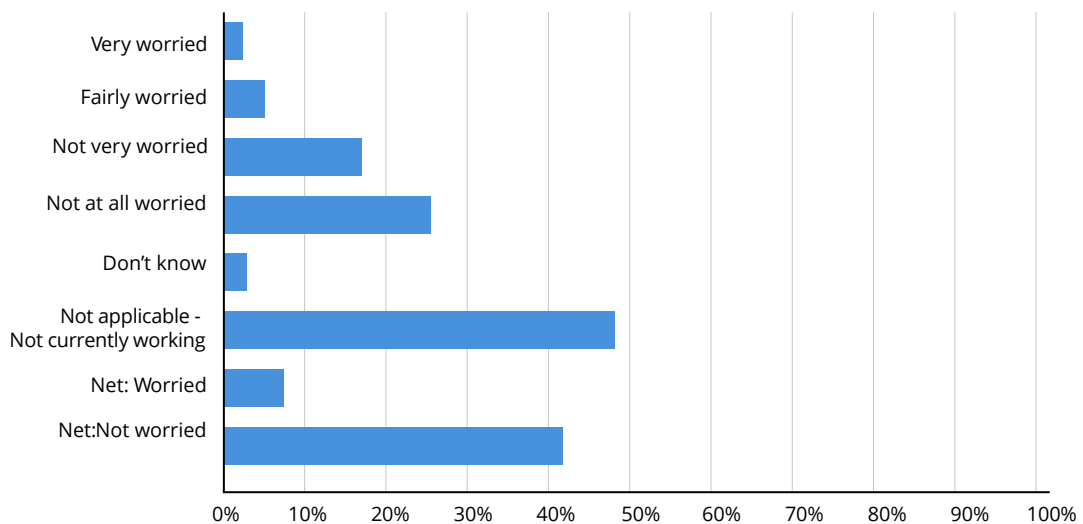
228. In this vein, it is also worth noting that much of the existing literature focuses on the technical potential for automating particular jobs or tasks, which does not necessarily translate into a risk that they will be automated in the real world.³⁰⁵ Many social and economic factors may influence whether a task is automated or not, beyond the technical potential to do so. What we do know is that employment by sector has changed dramatically over the past century. In 1901, manufacturing accounted for nearly 40% of employment across the country, and agriculture and fishing nearly 10%, but by 2011 these had fallen to 9% and 1% respectively, while the service economy now accounts for more than 80% of all employment (Figure 5). At least part of this change can be attributed to the impact of automation, and we would be remiss if we did not expect considerable change in employment patterns over the next 100 years.

304 Data for 1901 to 1911 covers Great Britain, while data for 1921 to 2011 covers only England and Wales. There is no census data for 1941, and the ONS did not include data for 1971 due to the difficulties of creating a consistent industrial grouping.

305 Written evidence from Future Advocacy (AIC0121)

229. A number of recent surveys suggest that the British public are significantly less concerned about automation affecting their own jobs than many experts are. A survey of 2108 adults conducted by YouGov, on behalf of Future Advocacy, between September and October 2017 found that the British public appeared to be comparatively unconcerned about the risks of losing their jobs to automation in the near future (as seen in Figure 6).

Figure 6: How worried are you that your job will be replaced by AI in the near future?



Total sample size was 2108 adults. Fieldwork was undertaken between 29th September and 2nd October 2017. The survey was carried out online. The figures have been weighted and are representative of all UK adults (aged 18+).

Source: Future Advocacy, *The impact of AI in UK constituencies: Where will automation hit hardest?* (October 2017), p 18: <https://static1.squarespace.com/static/5621e990e4b07de840c6ea69/t/59e777cd7bdce3041b57ac3/1508341775530/FutureAdvocacy-GeographicalAI.pdf> [accessed 7 February 2018]

230. A higher rate of concern was suggested by Demos, in a survey of 1234 adults in October 2017, which found that 35% thought there was “a risk [to their current jobs] from future developments in artificial intelligence and automation”, compared with 53% who believed there was no risk.³⁰⁶ However, CognitionX also cited a survey conducted by Arm and Northstar, which suggests that internationally concerns may be higher, with 57% of global respondents concerned that AI might become a risk to their jobs.³⁰⁷
231. **The labour market is changing, and further significant disruption to that market is expected as AI is adopted throughout the economy. As we move into this unknown territory, forecasts of AI’s growing impact—jobs lost, jobs enhanced and new jobs created—are inevitably speculative. There is an urgent need to analyse or assess, on an ongoing basis, the evolution of AI in the UK, and develop policy responses.**

306 Demos, ‘Public views on technology futures’ (29 November 2017): <https://www.demos.co.uk/project/public-views-on-technology-futures/> [accessed 1 February 2018]

307 Written evidence from CognitionX (AIC0170)

National Retraining Scheme

232. While the impact of AI on jobs remains highly uncertain, many of our witnesses believed that further Government assistance in terms of adult retraining, reskilling and lifelong learning would be an effective means preparation.³⁰⁸ In particular, the PHG Foundation argued that this should be focused on “skillsets that arguably cannot easily be displaced by AI such as creativity, effective social interaction, manual dexterity and intelligence”,³⁰⁹ while Research Councils UK suggested that “in-career re-skilling will become the norm every 10 years”.³¹⁰ Dr Ian Morgan and Brian Joyce also observed that, given that “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 ... financial support for those who wanted to retrain would be invaluable”.³¹¹
233. Future Advocacy highlighted the extent to which re-skilling could mitigate the impact of automation on jobs, providing the example of Accenture, where “17,000 jobs were automated but no-one lost their job, a feat that CEO of financial services Richard Lumb attributed to reskilling”.³¹²
234. There were notes of caution as well. Accenture emphasised the importance of ensuring that “those who were left behind by such fast-moving technological developments in the past: minorities, women, working mothers, disabled persons” needed to be included and prioritised in such efforts.³¹³ Professor Susskind said that consideration would need to be given to existing skillsets, as “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”.³¹⁴ In blue-collar jobs, he suggested this gap would likely be even bigger, and “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”. Future Advocacy highlighted the risks in this particular sector, noting that trials were planned for convoys of semi-automated lorries in the UK by the end of 2018, which posed a risk to the haulage and logistics industry’s 2.2 million employees.³¹⁵
235. For its part, the Government announced in its 2017 Autumn Budget that it would be establishing a National Retraining Scheme, aimed at helping people “re-skill and up-skill as the economy changes, including as a result of automation”.³¹⁶ The scheme will be guided by the National Retraining Partnership, which aims to bring together the Government, businesses and workers, through the CBI and the Trades Union Congress (TUC). It will initially focus on “priority skills”, with the first two named areas being digital and construction, funded with an initial investment of £64 million.³¹⁷ The scheme will be informed by the results of a £40 million programme to test

308 See for example Dr Toby Walsh ([AIC0078](#)); PHG Foundation ([AIC0092](#)); Amnesty International ([AIC0180](#)); Research Councils UK ([AIC0142](#)); Professor Richard Susskind ([AIC0194](#)); CognitionX ([AIC0170](#)) and IBM ([AIC0160](#))

309 Written evidence from PHG Foundation ([AIC0092](#))

310 Written evidence from Research Councils UK ([AIC0142](#))

311 Written evidence from Dr Ian Morgan and Brian Joyce ([AIC0179](#))

312 Written evidence from Future Advocacy ([AIC0121](#))

313 Written evidence from Accenture UK Limited ([AIC0191](#))

314 Written evidence from Professor Richard Susskind ([AIC0194](#))

315 Written evidence from Future Advocacy ([AIC0121](#))

316 *Industrial Strategy: Building a Britain fit for the future*, p 41

317 *Industrial Strategy: Building a Britain fit for the future*, p 11

“innovative approaches to helping adults up-skill and re-skill”, with particular emphasis on the use of AI and other innovative education technologies in online digital skills courses.³¹⁸ The Autumn Budget also announced an £8.5 million investment over the next two years in Unionlearn, a subsidiary of the TUC designed to boost learning in the workplace.³¹⁹

236. **The UK must be ready for the disruption that AI will have on the way in which we work. We support the Government’s interest in developing adult retraining schemes, as we believe that AI will disrupt a wide range of jobs over the coming decades, and both blue- and white-collar jobs which exist today will be put at risk. It will therefore be important to encourage and support workers as they move into the new jobs and professions we believe will be created as a result of new technologies, including AI. The National Retraining Scheme could play an important role here, and must ensure that the recipients of retraining schemes are representative of the wider population. Industry should assist in the financing of the National Retraining Scheme by matching Government funding. This partnership would help improve the number of people who can access the scheme and better identify the skills required. Such an approach must reflect the lessons learned from the execution of the Apprenticeship Levy.**

318 *Industrial Strategy: Building a Britain fit for the future*, p 117

319 *Autumn Budget 2017*, p 47

CHAPTER 6: LIVING WITH ARTIFICIAL INTELLIGENCE

237. The social implications of any new technology can often be overlooked in the excitement to embrace it. In this chapter we focus on the need to prepare future generations to engage and work with artificial intelligence, the potential affect it may have on social and political cohesion, and on inequality in the UK.

Education and artificial intelligence

238. Artificial intelligence, regardless of the pace of its development, will have an impact on future generations. The education system needs to ensure that it reflects the needs of the future, and prepares children for life with AI and for a labour market whose needs may well be unpredictable. Education in this context is important for two reasons. First, to improve technological understanding, enabling people to navigate an increasingly digital world, and inform the debate around how AI should, and should not, be used. Second, to ensure that the UK can capitalise on its position as a world leader in the development of AI, and grow this potential.

239. Our witnesses told us of the need to improve the data skills, digital understanding and literacy of young people in the UK. Google said: “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”.³²⁰ Baker McKenzie, a multinational law firm, told us that “education and training will be essential to prepare the workforce to use these emerging technologies effectively”.³²¹ In December 2017, the Federation of Small Businesses reported that 26% of small business owners lacked confidence in their basic digital skills and 22% believed that a lack of basic digital skills among their staff was preventing them from becoming more digital.³²² The Government recognised this too, and said they have “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”.³²³

240. Paul Clarke told us “what is not talked about enough is the fact that those skills lie at the end of what is a pipeline of digital literacy that stretches all the way back to primary school”.³²⁴ He said “I use the phrase ‘digital literacy’ as opposed to ‘coding’. I see digital literacy as being a much bigger portfolio. It includes things such as data literacy: how you harness data, how you visualise it, how you model it, how you understand bias”.³²⁵ Dr Mark Taylor, Global Strategy and Research Director for Dyson, agreed with Paul Clarke, and told us “it is extremely difficult to hire AI talent in the UK”.³²⁶

320 Written evidence from Google (AIC0225)

321 Written evidence from Baker McKenzie (AIC0111)

322 Federation of Small Businesses, *Learning the Ropes – Skills and training in small businesses* (11 December 2017), p 8: <http://www.fsb.org.uk/docs/default-source/fsb-org-uk/skills-and-training-report.pdf?sfvrsn=0> [accessed 25 January 2018]

323 Written evidence from HM Government (AIC0229)

324 Q 107 (Paul Clarke)

325 *Ibid.*

326 Q 107 (Dr Mark Taylor)

241. We heard that more emphasis should be placed on computer science in the overall curriculum. The UK Computing Research Committee, an Expert Panel of the British Computer Society, told us “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)”.³²⁷ The Committee also warned us that “initiatives to improve computing science education in the UK are poorly coordinated”.³²⁸ Dr Huma Shah and Professor Kevin Warwick, researchers in artificial intelligence, told us they “strongly believe that AI as a subject should be embedded into the school curriculum from primary age”.³²⁹ They explained that this could help increase the diversity of those working in AI.
242. Other witnesses pointed out that, with limited learning time, an increased focus on computer science will necessarily mean a reduction in other subjects. In particular they expressed concerns that the teaching of arts and humanities subjects which are closely linked with creative thinking, communication and the understanding of context, may suffer.³³⁰ Andrew Orłowski, Executive Editor of *The Register*, cautioned against the over-emphasis on computer science and literacy he believed was occurring in UK schools:
- “ ... my children go to an outstanding primary school in north London where they are taught algorithms every week but they are taught history once or twice a term and art, maybe, once or twice a term. There is an opportunity cost; there is only so much time for educating people. I question the value of teaching them algorithms. That is probably part of a balanced curriculum, but if they do not know culture and history how can they account for the world? ... You need those things probably more than you need to know how to use a computer”.³³¹
243. Miles Berry, Principal Lecturer, School of Education, University of Roehampton, agreed that “the breadth and balance of the curriculum is absolutely paramount”.³³² Future Advocacy said the Government’s reforms to technical education 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”.³³³
244. Others added nuance to this debate by suggesting the focus should be on digital understanding, rather than skills. Doteveryone said: “The best preparation the general public can have for AI, and indeed any technological change, is to have digital understanding”.³³⁴ They added “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”.³³⁵ Graham Brown-Martin, an education and technology researcher, told us that it was “far more important for young people to understand that the digital world is a built environment in exactly the same way the physical world is and that it contains all the biases and other limitations of the physical world” than

327 Written evidence from UK Computing Research Committee ([AIC0030](#))

328 *Ibid.*

329 Written evidence from Dr Huma Shah and Professor Kevin Warwick ([AIC0066](#))

330 [Q 39](#) (Dr Timothy Lanfear) and written evidence from Dr Jerry Fishenden ([AIC0028](#))

331 [Q 12](#) (Andrew Orłowski)

332 [Q 185](#) (Miles Berry)

333 Written evidence from Future Advocacy ([AIC0121](#))

334 Written evidence from Doteveryone ([AIC0148](#))

335 *Ibid.*

it is for them to be able to code.³³⁶ Professor Rosemary Luckin, Professor of Learner Centred Design at University College London, said “understanding the limitations of technology is really important, as is being able to demand from technology rather than being demanded of by technology”.³³⁷

245. We were told of the adverse effect an increasingly digital world was having on children in the UK. Professor Maja Pantic, Professor of Affective and Behavioural Computing, Imperial College London, said that children have reduced attention spans, shallower cognitive capabilities and experience a loss of identity as a result of time online and using social media.³³⁸ Professor Pantic warned us that the idealised world represented on social media “leads to many illnesses including eating disorders ... and serious mental illnesses”. Professor Pantic told us the increasing use of AI would add to this problem. Miles Berry told us the computing curriculum now requires schools to teach children from as early as five years old how to protect themselves in the digital world, for example, by keeping personal information private.³³⁹
246. After the Royal Society highlighted significant shortcomings in the National Curriculum’s approach to computer education in 2012,³⁴⁰ the Government introduced a new computing curriculum from September 2014,³⁴¹ aimed at addressing these problems and shifting education away from the use of basic software (the focus of ICT) towards coding and software development. Professor Hall told us that it was too early to tell what impact it was having.³⁴² Miles Berry helped design this new curriculum, and told us that the draft submitted to Ministers went “much further” than the current curriculum does in addressing personal morality in relation to technology. Berry said: “We included as an aim that children should be taught to develop an awareness of the individual and societal opportunities, challenges and risks raised by digital technology”.³⁴³
247. Professor Luckin said that establishing which ethics should be taught “would be a bit of a minefield” but “it is a conversation that we have to start having, because it is really important”.³⁴⁴ Graham Brown-Martin noted the absence of ethics in other digital arenas:

“At the moment within social media platforms we are seeing the results of not having ethics, which is potentially very damaging. You are talking about a question for society to answer in the public domain about what our ethics are. Just because we can do something does not mean that we should do it, and I think we are on that cusp”.³⁴⁵

336 [Q 184](#) (Graham Brown-Martin)

337 [Q 184](#) (Professor Rosemary Luckin)

338 Written evidence from Professor Maja Pantic ([AIC0215](#))

339 [Q 184](#) (Miles Berry)

340 Royal Society, *Shut down or restart: The way forward for computing in UK schools* (January 2012): <https://royalsociety.org/~media/education/computing-in-schools/2012-01-12-computing-in-schools.pdf> [accessed 23 January 2018]

341 Royal Society, *After the reboot: computing education in UK schools* (November 2017), p 17: <https://royalsociety.org/~media/policy/projects/computing-education/computing-education-report.pdf> [accessed 23 January 2018]

342 [Q 3](#) (Professor Dame Wendy Hall)

343 [Q 188](#) (Miles Berry)

344 [Q 188](#) (Professor Rosemary Luckin)

345 [Q 188](#) (Graham Brown-Martin)

248. On our visit to Cambridge, Microsoft Research told us that one of the central issues with computer science education at present was that it tended to be taught only, or primarily by, computer scientists.
249. **It is clear to us that there is a need to improve digital understanding and data literacy across society, as these are the foundations upon which knowledge about AI is built. This effort must be undertaken collaboratively by public sector organisations, civil society organisations (such as the Royal Society) and the private sector.**
250. **The evidence suggests that recent reforms to the computing curriculum are a significant improvement on the ICT curriculum, although it is still too early to say what the final results of this will be. The Government must be careful not to expand computing education at the expense of arts and humanities subjects, which hone the creative, contextual and analytical skills which will likely become more, not less, important in a world shaped by AI.**
251. **We are, however, concerned to learn of the absence of wider social and ethical implications from the computing curriculum, as originally proposed. We recommend that throughout the curriculum the wider social and ethical aspects of computer science and artificial intelligence need to be restored to the form originally proposed.**
252. Artificial intelligence, and more broadly, computer science, are fast-moving and complex areas to understand. Microsoft told us “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”.³⁴⁶ Other witnesses expressed similar views.³⁴⁷
253. In November 2017, the Royal Society published its report *After the reboot: computing education in UK schools*.³⁴⁸ This report found that there were no Teacher Subject Specialism Training courses available for computing; that in England the Government met only 68% of its recruitment target for new entrants to computing teacher training courses from 2012 to 2017; and that teachers felt the Government had changed the subject they teach, without providing them with sufficient support to teach it effectively. This confirms much of what our witnesses told us.
254. In February 2015 the House of Lords Select Committee on Digital Skills, in the summary to its report, *Make or Break: The UK’s Digital Future*, issued a robust call to arms which emphasised the need for urgency and cohesion in the delivery of the nation’s digital future.³⁴⁹ Three years later it is informative to revisit the well-considered recommendations of that Committee.
255. The Autumn Budget 2017 announced a number of measures aimed at improving computing education at the primary, secondary and further education stages. These included:

346 Written evidence from Microsoft (AIC0149)

347 See written evidence from Google (AIC0225); The Association for UK Interactive Entertainment (Ukie) (AIC0116) and Ocado Group plc (AIC0050)

348 The Royal Society, *After the Reboot – Computing Education in UK Schools* (November 2017): <https://royalsociety.org/~media/policy/projects/computing-education/computing-education-report.pdf> [accessed 31 January 2018]

349 Select Committee on Digital Skills, *Make or break: The UK’s Digital Future* (Report of Session 2014–15, HL Paper 111)

- Spending £84 million to train 8,000 computer science teachers (three times the number of existing teachers), with the aim that every secondary school has a fully qualified computer science GCSE teacher by the end of this Parliament.
- Establishing a National Centre for Computing with industry to produce training material and support schools with the teaching of computer science.
- Introducing measures to address the gender disparity between boys and girls studying STEM subjects at A-Level, as this hinders progress into higher education and careers in STEM.
- Providing additional funding for mathematics, including £27 million to be used to expand the Teaching for Mastery mathematics programme to 3,000 schools, and £40 million to establish Further Education Centres of Excellence across the country to train mathematics teachers and share best practice.

256. These steps are welcome, and look to address many of the immediate concerns of both the Royal Society and our own witnesses. However, until it is solved the issue will continue to be an acute shortage of confident well-trained, specialist teachers working across the public sector.

257. **While we welcome the measures announced in the Autumn Budget 2017 to increase the number of computer science teachers in secondary schools, a greater sense of urgency and commitment is needed from the Government if the UK is to meet the challenges presented by AI.**

258. **The Government must ensure that the National Centre for Computing is rapidly created and adequately resourced, and that there is support for the retraining of teachers with associated skills and subjects such as mathematics. In particular, Ofsted should ensure that schools are making additional time available to teachers to enable them to train in new technology-focused aspects of the curriculum. We also urge the Government to make maximum use across the country of existing lifelong learning facilities for the training and regular retraining of teachers and other AI experts.**

259. **Supplementary to the Hall-Pesenti Review, the Government should explore ways in which the education sector, at every level, can play a role in translating the benefits of AI into a more productive and equitable economy.**

Impact on social and political cohesion

260. AI may have many social and political impacts which extend well beyond people's lives as workers and consumers. The use of sophisticated data analytics for increasingly targeted political campaigns has attracted considerable attention in recent years, and a number of our witnesses were particularly concerned about the possible use of AI for turbo-charging this approach. The Leverhulme Centre for the Future of Intelligence highlighted the risk that "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”.³⁵⁰ Indeed, even the upbeat idea posed to us by John McNamara, a Master Inventor at IBM, of AI avatars trained on our “tastes, likes, dislikes [and] political views”, which could “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” raised troubling questions about the appropriate role for AI in mediating our democracy.³⁵¹

261. Witnesses also outlined the impact that AI might have on our wider perception of the world around us. The rise of ‘filter bubbles’—the idea that social media is increasingly feeding us information which aligns with our preconceived notions of the world, and closing us off from information which contradicts that world view—has been a much documented phenomenon. Witnesses said this phenomenon would likely be further exaggerated by AI. The Charities Aid Foundation (CAF) argued that “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)”.³⁵² This in turn could lead to “heightened social isolation and decreased community cohesion”. The BBC also expressed concerns that AI could “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”.³⁵³ The CAF also noted that the rise of ‘non-traditional interfaces’, such as conversational AI assistants, could heighten this effect by only providing one set of information in their responses.³⁵⁴
262. AI makes the processing and manipulating of all forms of digital data substantially easier and cheaper. Given that digital data permeates so many aspects of modern life, this presents opportunities, but also unprecedented challenges. AI could increasingly prove to have reality-distorting implications in other domains as well. In recent years researchers have shown how AI can be used to convincingly alter photographs and video footage, turning daylight scenes to night and placing words in the mouths of public figures.³⁵⁵ When we visited the BBC Blue Room we were shown an AI application which allows artificial copies of any individual’s voice to be replicated with relative ease.³⁵⁶ Witnesses also said a major challenge posed by AI was its potential use in the creation of fake news.³⁵⁷ Just as computer-generated imagery has transformed cinema and television in the past 20 years, we are now witnessing the emergence of AI applications which are allowing similar manipulations of everyday still and video footage and audio recordings on an industrial scale, without the need for extensive funding or expertise. This risks creating a world where nothing we see or hear can be taken on trust, and where ‘fake news’ becomes the default rather than the outlier.

350 Written evidence from Leverhulme Centre for the Future of Intelligence (AIC0182)

351 Written evidence from Mr John McNamara (AIC0081)

352 Written evidence from Charities Aid Foundation (AIC0042)

353 Written evidence from BBC (AIC0204)

354 Written evidence from Charities Aid Foundation (AIC0042)

355 James Vincent, ‘New AI research makes it easier to create fake footage of someone speaking’, *The Verge* (12 July 2017): <https://www.theverge.com/2017/7/12/15957844/ai-fake-video-audio-speech-obama> [accessed 1 February]; James Vincent, ‘NVIDIA uses AI to make it snow on streets that are always sunny’, *The Verge* (5 December 2017): <https://www.theverge.com/2017/12/5/16737260/ai-image-translation-nvidia-data-self-driving-cars> [accessed 1 February 2018]

356 See, for examples of this, <https://lyrebird.ai/>

357 Written evidence from Accenture UK Limited (AIC0191); Future Intelligence (AIC0216) and Q 24 (Dr Ing Konstantinos Karachalios)

263. Our witnesses also expressed concern that if too many political decisions are delegated to machines, the feelings of powerlessness and exclusion felt by some could be further amplified. As Future Intelligence said, “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”.³⁵⁸ Dr Andrew Blick, Senior Lecturer in Politics and Contemporary History at King’s College London, suggested to us that AI might have a profound impact on the way that decisions are made by Government Ministers, the advice given by and to the civil service. He added that “if artificial intelligence can lead to the more effective delivery of services required by the public, it is desirable from a democratic perspective” but could challenge the concept of ministerial responsibility to Parliament.³⁵⁹
264. On 23 January 2018, the Government announced that it was establishing a National Security Communications Unit within the Cabinet Office “tasked with combating disinformation by state actors and others”.³⁶⁰ The role of the unit, and the impact of this approach, remains to be seen.
265. **There are many social and political impacts which AI may have, quite aside from people’s lives as workers and consumers. AI makes the processing and manipulating of all forms of digital data substantially easier, and given that digital data permeates so many aspects of modern life, this presents both opportunities and unprecedented challenges. As discussed earlier in our report, there is a rapidly growing need for public understanding of, and engagement with, AI to develop alongside the technology itself. The manipulation of data in particular will be a key area for public understanding and discussion in the coming months and years.**
266. **We recommend that the Government and Ofcom commission research into the possible impact of AI on conventional and social media outlets, and investigate measures which might counteract the use of AI to mislead or distort public opinion as a matter of urgency.**

Inequality

267. The growing prevalence of AI raises questions about how economic inequality will be addressed in future. Some economists, most notably David Autor, have argued that the polarisation of the job market over the past thirty years, towards low-skilled and high-skilled jobs, and away from medium-skilled jobs, is likely to be reversed, as some low- and medium-skilled jobs are likely to be relatively resistant to automation, while some highly-skilled but relatively routine jobs may be automatable with AI.³⁶¹ But most agree that automation is likely to mean that highly-skilled workers, who are typically more adaptable and will have a larger stake in AI, are likely to take a growing proportion of income, while low-skilled workers, who have typically struggled to adapt to technological change and will have at least some work taken away from them by machines, are more likely to struggle.

358 Written evidence from Future Intelligence (AIC0216)

359 Written evidence from Dr Andrew Blick (AIC0064)

360 ‘Government announces anti-fake news unit’, *BBC News* (23 January 2018): <http://www.bbc.co.uk/news/uk-politics-42791218> [accessed 24 January 2018]

361 David Autor, ‘Why are there still so many jobs? The history and future of workplace automation’ in *The Journal of Economic Perspectives*, vol. 29, no. 3 (2015): <https://economics.mit.edu/files/11563> [accessed 1 March 2018]

268. The Charities Aid Foundation echoed the concerns of many when they told us that 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”.³⁶² Research Councils UK emphasised that “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”, while the BBC noted that “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”.³⁶³ Sarah O’Connor said:

“The big question that people in the economics and labour market world are thinking about is: how will those gains be distributed? If indeed AI leads to vast increases in efficiency, using fewer workers, does that mean that all the wealth that is created from that will go to the people who own the AI—the intellectual property—and the data that feeds into it? If so, what does that mean for the people who might be displaced out of jobs? Will there be new jobs to replace those old ones? If there are, will they be of the same quality?”³⁶⁴

269. Olly Buston said that, given the nature of jobs most amenable to automation, inequality may develop unevenly across different parts of the country, with most parts of London potentially faring well, and parts of the Midlands and the north of England suffering the most.³⁶⁵ Sarah O’Connor recently emphasised the need to think about automation-related inequality in terms of places, as people are often far less mobile than economists might like, and without new jobs in smaller towns and more deprived parts of the UK, regional inequality will prove very difficult to tackle.³⁶⁶

270. A range of approaches have been suggested, which fall into two broad categories. The first is to focus on re-training, as the Government appears to be doing with its recent announcement of a National Retraining Scheme, as discussed in Chapter 5.

271. The second is to pursue more radical policies for redistributing the gains from AI and automation. Of these, the most discussed is the concept of a ‘universal basic income’ (UBI), whereby everyone would be provided with a standardised monthly income from the Government. This would replace most if not all other forms of welfare payment, and would be paid regardless of whether people were in work or not. We received a range of opinions on this subject. Many of our witnesses expressed interest in UBI, and were supportive of pilot schemes being carried out in various parts of the world, with Scotland being the most recent example.³⁶⁷ However, a number of reservations were also expressed, with some witnesses believing it was

362 Written evidence from Charities Aid Foundation ([AIC0042](#))

363 Written evidence from Research Councils UK ([AIC0142](#)) and BBC ([AIC0204](#))

364 [Q 10](#) (Sarah O’Connor)

365 [Q 97](#) (Olly Buston)

366 Sarah O’Connor, ‘Our robot era demands a different approach to retraining’, *Financial Times* (23 January 2018): <https://www.ft.com/content/c4bde676-0027-11e8-9650-9c0ad2d7c5b5> [accessed 24 January 2018]

367 Written evidence from Dr Andrew Pardoe ([AIC0020](#)); Deep Learning Partnership ([AIC0027](#)); 10x Future Technology ([AIC0024](#)); Future Advocacy ([AIC0121](#)); IEEE European Public Policy Initiative Working Group on ICT ([AIC0106](#)); Mr Thomas Cheney ([AIC0098](#)) and Amnesty International ([AIC0180](#))

premature to consider such a radical measure,³⁶⁸ while others argued that work provided people with a sense of meaning and purpose, which could not be addressed with a simple cash payment.³⁶⁹

272. Future Advocacy raised the idea of a so-called ‘robot tax’, most prominently suggested by Bill Gates, on companies which adopt automating technologies, in order to support redistributive and retraining initiatives. They noted that it might “provide a solution to the potential problem that reduced employment will lead to reduced income tax and National Insurance revenues”, which together account for almost 60% of total tax revenue. However, they also told us that more work needed to be done to ensure such an idea would foster rather than hinder innovation.³⁷⁰
273. The Government’s response to this question has so far been mixed at best. As discussed in Chapter 5, the National Retraining Scheme is a promising initiative, although it will require a considerable political commitment to ensure its success. In terms of the potential for regional inequality, we are pleased that the Government’s Industrial Strategy discussed regional development at length, and plans for local industrial strategies, which will complement the national strategy, are to be applauded.³⁷¹ On the other hand, the Government has not explicitly engaged with the possibility of AI and automation-related inequality, either in its Industrial Strategy or its response to our call for evidence. Meanwhile, the early interest taken by the Prime Minister in improving social mobility appears to have been deprioritised, and the recent resignation of all four members of the Social Mobility Commission, citing a lack of progress, does not bode well.³⁷² More recently, the Prime Minister stated at Davos that, alongside the “opportunities of technology” such as AI, there was a need to “shape this change to ensure it works for everyone”, but it remains to be seen whether these sentiments are followed up with concrete action.
274. The Scottish Government has taken a different tack, and are supporting four areas in Scotland—Glasgow, Fife, Edinburgh and North Ayrshire—in the design of pilot basic income schemes, with £100,000 allocated to the schemes in the draft budget, and additional funds coming from local budgets.³⁷³ The civil service has estimated that a Scotland-wide roll-out would cost around £12.3 billion. Nicola Sturgeon MSP, First Minister of Scotland, said: “It might turn out not to be the answer, it might turn out not to be feasible. But as work and employment changes as rapidly as it is doing, I think it’s really important that we are prepared to be open-minded about the different ways that we can support individuals to participate fully in the new economy”.³⁷⁴ However, it is expected that it will take between 12 and 18 months to design the schemes, and it seems likely that it will take several years after that before any results are forthcoming.³⁷⁵

368 Written evidence from Sage (AIC0159) and Q 11 (Sarah O’Connor)

369 Written evidence from Dr Paula Boddington (AIC0067); The Knowledge, Skills and Experience Foundation (AIC0044); Q 98 (Olly Buston) and Charities Aid Foundation (AIC0042)

370 Written evidence from Future Advocacy (AIC0121)

371 *Industrial Strategy: Building a Britain fit for the future*, pp 214–239

372 Jim Pickard, ‘Theresa May’s Social Mobility Commission walks out’, *Financial Times* (3 December 2017): <https://www.ft.com/content/e4426dce-d808-11e7-a039-c64b1c09b482> [accessed 22 January 2018]

373 Philip Sim, ‘Citizen’s income: Could it work in Scotland?’, *BBC News* (27 December 2017): <http://www.bbc.co.uk/news/uk-scotland-scotland-politics-41832065> [accessed 22 January 2018]

374 Libby Brooks, ‘Scotland united in curiosity as councils trial universal basic income’, *The Guardian* (25 December 2017): <https://www.theguardian.com/uk-news/2017/dec/25/scotland-universal-basic-income-councils-pilot-scheme> [accessed 22 January 2018]

375 Philip Sim, ‘Citizen’s income: Could it work in Scotland?’, *BBC News* (27 December 2017): <http://www.bbc.co.uk/news/uk-scotland-scotland-politics-41832065> [accessed 22 January 2018]

275. **The risk of greater societal and regional inequalities emerging as a consequence of the adoption of AI and advances in automation is very real, and while the Government's proposed policies on regional development are to be welcomed, we believe more needs to be done in this area. We are not yet convinced that basic income schemes will prove to be the answer, but we watch Scotland's experiments with interest.**
276. **Everyone must have access to the opportunities provided by AI. The Government must outline its plans to tackle any potential societal or regional inequality caused by AI, and this must be explicitly addressed as part of the implementation of the Industrial Strategy. The Social Mobility Commission's annual State of the Nation report should include the potential impact of AI and automation on inequality.**

CHAPTER 7: HEALTHCARE AND ARTIFICIAL INTELLIGENCE

277. This chapter is a case study on the use of artificial intelligence in the healthcare sector in the UK. Many of the issues presented by AI when deployed in healthcare are representative of wider issues with the use of artificial intelligence, such as the possible benefits to individuals and for the public good, the handling of personal data, public trust, and the need to mitigate potential risks.

The opportunity

278. Our witnesses were clear that healthcare was one sector where AI presented significant opportunities. The Academy of Medical Science said “the impact of artificial intelligence on ... the healthcare system is likely to be profound” because research and development will become more efficient, new methods of healthcare delivery will become possible, clinical decision-making will be more informed, and patients will be more informed in managing their health.³⁷⁶ Others agreed with this assessment.³⁷⁷ Professor John Fox, who has worked in the field of AI for healthcare for over three decades, sounded a rare dissenting note, and suggested that many of the claims for healthcare AI may well be overblown. Professor Fox said there was a need for Parliament to commission a “dispassionate and objective study of the evidence for success in healthcare” due to the level of interest and a lack of critical analysis of developments.³⁷⁸

279. Some of our witnesses told us of the specific areas within healthcare that could benefit from artificial intelligence. The Royal College of Radiologists told us “medical imaging is perfectly placed to benefit from advances in AI” because of the availability of high quality, curated data, “overcoming one of the main hurdles in AI development”.³⁷⁹ At Microsoft Research in Cambridge, we saw their work on ‘InnerEye’ technology, which is being developed to assist oncologists in the analysis of x-ray and MRI scans. Such an application has the potential to dramatically reduce the cost of analysing scans, allowing far more to be taken over the course of a treatment, thereby facilitating more accurately targeted treatment. The Royal College of Radiologists told of the potential for more efficient breast imaging, where one of two human breast screen readers could be replaced with AI (as mammograms are conventionally double read by a radiologist, advanced practitioner or breast physician). With two million women screened every year in the UK, and with images read at a rate of 55 per hour, considerable time could be saved for the specialists involved. The Royal College told us this was of particular importance given the strain the service is under due to staffing shortages and that many consultants were due to retire at the same time, 30 years after the breast screening programme was established.³⁸⁰ The Medicines and Healthcare products Regulatory Agency (MCHR) gave us a detailed list of possible uses of AI in healthcare, including for genomics and personalised medicine, the detection and monitoring of pandemics or epidemics, and the development of evidence for medicines submissions.³⁸¹

376 Written evidence from the Academy of Medical Sciences ([AIC0210](#))

377 Written evidence from the Wellcome Trust and Association of Medical Research Charities ([AIC0202](#)) and [Q 2](#) (Professor Dame Wendy Hall)

378 Written evidence from Professor John Fox ([AIC0076](#))

379 Written evidence from Royal College of Radiologists ([AIC0146](#))

380 *Ibid.*

381 Written evidence from Medicines and Healthcare products Regulatory Agency ([AIC0134](#))

280. Other witnesses pointed to the more administrative benefits AI could offer the NHS. Deloitte said “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”.³⁸² Braintree, an artificial intelligence research and development company, said: “instead of being buried in administrative duties and routine medical analysis, [doctors] could concentrate more fully on patient care and higher level medical diagnosis”.³⁸³ When we visited DeepMind they told us of their work with Moorfields Eye Hospital, which they hope will reduce the delay between patients being seen and then treated for certain eye conditions. Touch Surgery, a platform which offers training for surgeons, said AI could also help to “inform scheduling systems, updating them with real-time sensor feeds, to better utilise resources and availability”.³⁸⁴
281. However, the Centre for Health Economics at the University of York warned that “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”. The Centre suggested that a decision needed to be made as to whether devoting scarce resource to the adoption of AI was appropriate, and the Government needed to provide a clear answer to this question.³⁸⁵
282. It is no secret that the NHS is under immense pressure. For any of the benefits outlined above to be realised, the use of AI in healthcare is dependent on a number of factors, including:
- the acceptance by the public of AI playing a role in their treatment;
 - the use of patient data;
 - the NHS being equipped to deploy new technology; and
 - staff trained in how to use it.

The value of data

283. The NHS holds data on nearly everyone in the UK; some of it going back decades. Lord Henley recognised this intrinsic value when he told us the “advantage of having a National Health Service is the quantity and the quality of the data that we have, which other countries do not necessarily have”.³⁸⁶
284. Our witnesses agreed that this data could be of immense value to artificial intelligence researchers. Dr Hugh Harvey, a consultant radiologist and artificial intelligence researcher, suggested that IBM’s acquisition of Merge Healthcare in the USA for \$1 billion, which netted them five to six million patients’ records, might be indicative of the value of the data held by the NHS. He also pointed to the Royal Society’s report, *Machine Learning: the power and promise of computers that learn by example*, which cited a figure of

382 Written evidence from Deloitte ([AIC0075](#))

383 Written evidence from Braintree ([AIC0074](#))

384 Written evidence from Touch Surgery ([AIC0070](#))

385 Written evidence from the Centre for Health Economics, University of York ([AIC0242](#))

386 [Q 192](#) (Lord Henley)

£1.8 billion for the direct value of public sector data, and put the wider socio-economic benefits at a minimum of £6.8 billion.³⁸⁷

285. Nicola Perrin, who leads the Understanding Patient Data initiative at the Wellcome Trust, said that the question of ascertaining the value of the data the NHS holds, and the nature of compensation that should be sought for access to any data was “a crucial one to get right because of the implications for public confidence”. She added that the public “do not like the idea of the NHS selling data, but they are even more concerned if companies are making a profit at the expense of both the NHS and patients”.³⁸⁸
286. Some argued that compensation for access to data does not necessarily have to be financial: it is clear that by sharing data with researchers, it would be fair for the NHS to expect favourable (if not free) access to any AI-based products developed from patient data. DeepMind agreed “that the NHS should be recompensed when it makes data available to companies for the purposes of AI development”, but said “clearly there are many ways to recognise and return value”.³⁸⁹ With the development of Streams, although not an AI application, DeepMind have given the Royal Free London NHS Foundation Trust “five years’ free use of the system” in exchange for testing the application (see Box 8).³⁹⁰
287. Dr Harvey said there should not be a “monetary barrier to entry to data access”, and that “we need to encourage innovation and have failure, and they need to be allowed to fail at low cost”.³⁹¹ Professor Martin Severs, Medical Director for NHS Digital (the national information and technology partner to the health and social care system), said:
- “I would not have a barrier for entry but I would have some mechanism of demonstrating societal benefit from the data as it is being used. NHS Digital is open to any of those which have a consistent buy-in by all the organisations”.³⁹²
288. Professor Severs thought that the public would see the use of their data to develop an AI tool as a “fair deal” if it had a wider societal benefit.³⁹³ Dr Barber told us of the “concept called the ‘golden share’ which enables companies to give money back to the contributors of the data” which could be applied to arrangements with the NHS and AI researchers.³⁹⁴ He said that the Department for Transport was already using such an arrangement.
289. Other witnesses, however, told us it may be hard to capitalise on the value of the data. Dr Julian Huppert, Chair of the Independent Review Panel for DeepMind Health, said “the public tend to believe that the NHS is one institute which has all the data in one place”. He said there are “real problems with data storage, availability and flow throughout the NHS at pretty much every level. It is very much in silos at the moment.” DeepMind told us that “the

387 [Q 132](#) (Dr Hugh Harvey); The Royal Society, *Machine Learning: the power and promise of computers that learn by example* (April 2017): <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf> [accessed 16 January 2018]

388 [Q 120](#) (Nicola Perrin)

389 Written evidence from DeepMind ([AIC0234](#))

390 [Q 120](#) (Dr Julian Huppert)

391 [Q 132](#) (Dr Hugh Harvey)

392 [Q 132](#) (Professor Martin Severs)

393 *Ibid.*

394 [Q 11](#) (Dr David Barber)

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”.³⁹⁵ Dr Harvey told us:

“Medical data ... is very chaotic at source. This comes down to a delay, specifically in the NHS but also across the world, in the technology that is available in healthcare institutions compared to the technology that is available on the high street”.³⁹⁶

290. Dr Bunz and Elizabeth Denham, the Information Commissioner, reminded us, as we have discussed earlier, that data is not necessarily owned but rather controlled: each time the data is processed, value can be added.³⁹⁷ Put simply, an individual’s data might be worth very little on its own. When that data is brought together into a dataset in an NHS database, its value increases. When work is done to prepare that dataset for training an algorithm, the data is worth even more. At each stage, the value grows.
291. Another complication in the assessment of the value of data, and the sort of compensation that ought to be expected, is the piecemeal arrangements being made between NHS trusts and companies, some of which may have far more experienced negotiators than trusts have access to.

Box 8: DeepMind and the Royal Free London NHS Foundation Trust

DeepMind is an artificial intelligence research company, based in London, and owned by Alphabet. In 2015, DeepMind began working with the Royal Free London NHS Foundation Trust to develop an app to help with the diagnosis of acute kidney injury (AKI). Subsequently, the Streams app was developed and deployed within the Trust.

In the development of Streams, the Trust provided personal data of around 1.6 million patients as part of a trial to test an alert, diagnosis and detection system for AKI. When this came to light, the Information Commissioner’s Office (ICO) investigated, and ruled that the Trust failed to comply with the Data Protection Act 1998 when it provided patient details to DeepMind. Although the app does not use artificial intelligence or deep learning techniques, DeepMind’s involvement has highlighted some of the potential issues involved in using patient data to develop technological solutions, many of which are relevant to AI.

Source: Information Commissioner’s Office, ‘Royal Free—Google DeepMind trial failed to comply with data protection law’: <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/> [accessed 8 February 2018]

292. With this, it is important to bear in mind the lessons of, as one witness described it, the “Royal Free Hospital/DeepMind fiasco”.³⁹⁸ Doteveryone said this case exemplified what were “many of the major issues at 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”.³⁹⁹ Nicola Perrin told us that it was “absolutely the situation” that NHS trusts were separately (more or less entrepreneurially) making different arrangements with different companies to use datasets that

395 Written evidence from DeepMind (AIC0234)

396 Q 131 (Dr Hugh Harvey)

397 Q 56 (Dr Mercedes Bunz and Elizabeth Denham)

398 Written evidence from medConfidential (AIC0063)

399 Written evidence from Doteveryone (AIC0148)

are very variable in their worth, suitability and application.⁴⁰⁰ Dr Huppert said “there are lots of different providers in lots of different trusts and the system is very chaotic” and that “there has not been very much work to look at some of the providers whose standards are not very high”.⁴⁰¹

293. This lack of consistency not only risks the NHS not maximising the value of data it holds, but also risks the sharing of intensely personal patient data. The sharing of data, even with the best of intentions, to companies which may not be equipped to handle such data securely, must be avoided at all costs.

Box 9: Caldicott Guardians

A Caldicott Guardian is a senior official responsible for protecting the confidentiality of people’s health and care information and enabling appropriate information-sharing. All NHS organisations (since 1999) and local authorities which provide social services (since 2002) must have a Caldicott Guardian. The Guardian plays a role in ensuring that their organisation satisfies the highest standards for handling patient identifiable information, and advises on options for the lawful and ethical processing of information. The role has no statutory basis, and is mostly advisory in nature: however, the Guardian is accountable for any advice given.

There are seven Caldicott Principles which guide the advice of a Guardian. These are:

- justify the purpose;
- do not use personal confidential data unless absolutely necessary;
- use the minimum necessary personal confidential data;
- access to personal confidential data should be on a strict need-to-know basis;
- everyone with access to such data should be aware of their responsibilities;
- comply with the law; and
- the duty to share information can be as important as the duty to protect patient confidentiality.⁴⁰²

Source: HM Government, ‘UK Caldicott Guardian Council’: <https://www.gov.uk/government/groups/uk-caldicott-guardian-council> [accessed 8 February 2018]

294. Dame Fiona Caldicott, the National Data Guardian, described the challenge of using patient data in technology, and its implications:

“What we have not done is take the public with us in these discussions, and we really need their views. What is the value? Are they happy for their data to be used when it is anonymised for the purposes we have described? We need to have the public with us on it, otherwise they will be upset that they do not know what is happening to their data and be unwilling to share it with the people to whom they turn for care. That is the last thing we want to happen in our health service”.⁴⁰³

400 Q 120 (Nicola Perrin)

401 Q 120 (Dr Julian Huppert)

402 UK Caldicott Guardian Council, *A manual for Caldicott Guardians* (2017): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/581213/cgmanual.pdf [accessed 1 February 2018]

403 Q 132 (Dame Fiona Caldicott)

295. The patchwork approach is not just a challenge for the NHS. Perrin said “from a company perspective, it is very difficult for them to know how to access the NHS which is a big beast and some hospitals have much easier conversations than others”.⁴⁰⁴ Dr Sobia Raza, Head of Science, PHG Foundation, advocated a more joined up approach, which could help in “realising the benefits in terms of negotiations with companies and developing a dataset that could provide more opportunities for accurate tools and algorithms”.⁴⁰⁵ Dr Raza also spoke to the benefits of having access to data at a national level, instead of at a local one: “a huge opportunity arises when you can capture the differences in demographics and, essentially, collate a more enriched dataset which is more reflective of the wider population”.⁴⁰⁶ We were encouraged by Professor Martin Severs, who told us “NHS Digital would support a national, open and consistent approach”.⁴⁰⁷

Using AI

296. We were concerned by the NHS’s lack of organisational preparedness to embrace new technology. In April 2017, the Select Committee on the Long-Term Sustainability of the NHS concluded “there is a worrying absence of a credible strategy to encourage the uptake of innovation and technology at scale across the NHS”.⁴⁰⁸ In July 2017, The DeepMind Health Independent Review Panel Annual Report stated “the digital revolution has largely bypassed the NHS, which, in 2017, still retains the dubious title of being the world’s largest purchaser of fax machines”.⁴⁰⁹ Dr Huppert, Chair of the Panel, told us that “there is a huge amount of work that is still needed to make the NHS more digitally savvy”.⁴¹⁰

297. When asked if the NHS had the capacity to take advantage of the opportunities, and to minimise the risks, of using AI, Dr Huppert said “the short answer is no”, although “clinicians vary: some of them are very technologically savvy and very keen and eager and some of them very much are not”.⁴¹¹ Dr Raza said “there is an important need here for healthcare professionals to have knowledge about the technology, to be aware of what it is capable of and to understand its limitations and gauge an awareness of how it might change or influence clinical practice in years to come”.⁴¹² Dr Raza also told us it was just as important that those developing AI engaged with healthcare professionals at an early stage to help them establish where artificial intelligence could be of the most use. Dr Raza said there needed to be “a continued drive towards digitisation and embedding appropriate and modern digital infrastructure” in the NHS.⁴¹³ Nicola Perrin highlighted the establishment of the NHS Digital Academy, a virtual organisation which provides a year-long digital health training course for Chief Clinical Information Officers, Chief Information Officers, and those interested within the NHS from clinical or non-clinical backgrounds.

404 [Q 120](#) (Nicola Perrin)

405 [Q 120](#) (Dr Sobia Raza)

406 *Ibid.*

407 [Q 132](#) (Professor Martin Severs)

408 Select Committee on the Long-term Sustainability of the NHS, *The Long-Term Sustainability of the NHS and Adult Social Care* (Report of Session 2016–17, HL Paper 151)

409 DeepMind, ‘Independent Reviewers release first annual report on DeepMind Health’ (5 July 2017): <https://deepmind.com/blog/independent-reviewers-annual-report-2017/> [accessed 18 January 2018]

410 [Q 120](#) (Dr Julian Huppert)

411 *Ibid.*

412 [Q 122](#) (Dr Sobia Raza)

413 [Q 122](#) (Dr Sobia Raza) and [Q 127](#) (Dr Sobia Raza)

298. Professor Severs told us that clinicians would embrace technology if it could help alleviate them of time-consuming, routine tasks.⁴¹⁴ Both Dame Fiona Caldicott and Dr Harvey agreed that a multidisciplinary approach was required.⁴¹⁵ Dr Harvey said “the medical syllabus needs to start incorporating not just medical statistics but some basics of data science”, as the NHS could not compete with the high salaries offered by industry to dedicated medical data scientists and researchers. Dr Harvey suggested collaboration between everyone with an interest in the healthcare system and AI was needed, as “if the NHS was to try to do it on its own it would fall short of the relevant skills and funding to do so”.⁴¹⁶ It is clear to us that more needs to be done to ensure that everyone in the NHS is equipped to embrace the potential of AI in healthcare, identify and minimise the possible risks.
299. The application of artificial intelligence in the delivery of healthcare in the UK offers significant opportunities to improve the diagnosis and treatment of the unwell, as well as to help the NHS and other healthcare providers be more efficient. Further research and innovation should be encouraged by the NHS, and by the Government, while the public must be reassured that their data will not be made available for use without appropriate safeguards in place. The NHS should look to assess where the most value can be gained from the use of AI in the delivery of its services, and where the patient experience can be most improved through its deployment.
300. **Maintaining public trust over the safe and secure use of their data is paramount to the successful widespread deployment of AI and there is no better exemplar of this than personal health data. There must be no repeat of the controversy which arose between the Royal Free London NHS Foundation Trust and DeepMind. If there is, the benefits of deploying AI in the NHS will not be adopted or its benefits realised, and innovation could be stifled.**
301. **The data held by the NHS could be considered a unique source of value for the nation. It should not be shared lightly, but when it is, it should be done in a manner which allows for that value to be recouped. We are concerned that the current piecemeal approach taken by NHS Trusts, whereby local deals are struck between AI developers and hospitals, risks the inadvertent under-appreciation of the data. It also risks NHS Trusts exposing themselves to inadequate data sharing arrangements.**
302. **We recommend that a framework for the sharing of NHS data should be prepared and published by the end of 2018 by NHS England (specifically NHS Digital) and the National Data Guardian for Health and Care. This should be prepared with the support of the ICO and the clinicians and NHS Trusts which already have experience of such arrangements (such as the Royal Free London and Moorfields Eye Hospital NHS Foundation Trusts), as well as the Caldicott Guardians. This framework should set out clearly the considerations needed when sharing patient data in an appropriately anonymised form, the precautions needed when doing so, and an awareness of the value of that data and how it is used. It must also take account of the need to ensure SME access to NHS data, and ensure that patients are made aware of the use of their data and given the option to opt out.**

414 [Q 138](#) (Professor Martin Severs)

415 [Q 138](#) (Dame Fiona Caldicott; Dr Hugh Harvey)

416 [Q 138](#) (Dr Hugh Harvey)

303. **Many organisations in the United Kingdom are not taking advantage of existing technology, let alone ready to take advantage of new technology such as artificial intelligence. The NHS is, perhaps, the most pressing example of this. The development, and eventual deployment, of AI systems in healthcare in the UK should be seen as a collaborative effort with both the NHS and the AI developer being able to benefit. To release the value of the data held, we urge the NHS to digitise its current practices and records, in consistent formats, by 2022 to ensure that the data it holds does not remain inaccessible and the possible benefits to society unrealised.**

CHAPTER 8: MITIGATING THE RISKS OF ARTIFICIAL INTELLIGENCE

304. In the course of our inquiry we encountered several serious issues associated with the use of artificial intelligence that require careful thought, and deliberate policy, from the Government. These include the issue of determining legal liability, in cases where a decision taken by an algorithm has an adverse impact on someone's life, the potential criminal misuse of artificial intelligence and data, and the use of AI in autonomous weapons systems.

Legal liability

305. The emergence of any new technology presents a challenge for the existing legal and regulatory framework. This challenge may be made most apparent by the widespread development and use of artificial intelligence. Cooley (UK) LLP told us that “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)”.⁴¹⁷

306. A serious issue which witnesses brought to our attention was who should be held accountable for decisions made or informed by artificial intelligence. This could be a decision about receiving a mortgage, in diagnosing illness, or a decision taken by an automated vehicle on the road.

307. Arm, a multinational semiconductor and software design company, asked: “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”.⁴¹⁸ The IEEE's European Public Policy Initiative Working Group on ICT told us that one of the major legal issues which needed to be addressed was the establishment of “liability of industry for accidents involving autonomous machines” because “this poses a challenge to existing liability rules where a legal entity (person or company) is ultimately responsible when something goes wrong”.⁴¹⁹

308. Our witnesses explained why addressing the question of legal liability was so important. The Royal College of Radiologists said “legal liability is often stated as a major societal hurdle to overcome before widespread adoption of AI becomes a reality”.⁴²⁰ Dr Mike Lynch said a legal liability framework and insurance were “vital to allow these systems to actually be used. If insurance and legal liability are not sorted out this will be a great hindrance to the technology being adopted”.⁴²¹ We agree with our witnesses in this regard. Unless a clear understanding of the legal liability framework is reached, and steps taken to adjust such a framework if proven necessary, it is foreseeable that both businesses and the wider public will not want to use AI-powered tools.

309. Our witnesses considered whether, in the event AI systems malfunction, underperform or otherwise make erroneous decisions that cause individuals harm, new mechanisms for legal liability and redress in these situations were

417 Written evidence from Cooley (UK) LLP ([AIC0217](#))

418 Written evidence from Arm ([AIC0083](#))

419 Written evidence from the IEEE European Public Policy Initiative Working Group on ICT ([AIC0106](#))

420 Written evidence from Royal College of Radiologists ([AIC0146](#))

421 Written evidence from Dr Mike Lynch ([AIC0005](#))

needed. Kemp Little LLP told us that our current legal system looks to establish liability based on standards of behaviour that could be reasonably expected, and looks to establish the scope of liability based on the foreseeability of an outcome from an event.⁴²² They told us “AI challenges both of these concepts in a fundamental way”.⁴²³ This is because of the difficulties which exist in understanding how a decision has been arrived at by an AI system. Kemp Little LLP also suggested that “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”.⁴²⁴ In contrast, Professor Reed thought the existing legal mechanisms worked: “The law will find a solution. If we have a liability claim, the law will find somebody liable or not liable”.⁴²⁵ Professor Reed did, however, also tell us that some of the questions asked to identify liability “may be answerable only by obtaining information from the designers who are from a different country. It will make litigation horribly expensive, slow, and very difficult”.⁴²⁶

310. Professor Karen Yeung, then Professor of Law and Director of the Centre for Technology, Ethics, Law and Society, Dickson Poon School of Law, King’s College London, said that she did “not think that our existing conceptions of the liability and responsibility have yet adapted” and “that if it comes to court the courts will have to find a solution, but somebody will have been harmed already”.⁴²⁷ Professor Yeung told us “it is in the interests of industry and the general public to clarify and provide assurance that individuals will not suffer harm and not be uncompensated”.⁴²⁸ Paul Clarke, Ocado, said:

“AI definitely raises all sorts of new questions to do with accountability. Is it the person or people who provided the data who are accountable, the person who built the AI, the person who validated it, the company which operates it? I am sure much time will be taken up in courts deciding on a case-by-case basis until legal precedence is established. It is not clear. In this area this is definitely a new world, and we are going to have to come up with some new answers regarding accountability”.⁴²⁹

311. Others from industry did not agree. Dr Mark Taylor, Dyson, told us that he does “not foresee a situation with our products where we would fall outside the existing legislation” as anything Dyson sells complies with the laws and regulations of the market in which they sell them.⁴³⁰ Dr Joseph Reger, Chief Technology Officer for Europe, the Middle East, India and Africa at Fujitsu, told us that “we need a legal system that keeps up ... because these products are hitting the market already and therefore the questions of liability, responsibility and accountability need to have a new definition very soon”.⁴³¹ It is clear to us, therefore, that the issue of liability needs to be addressed as soon as possible, in order to ensure that it is neither a barrier to widespread adoption, nor decided too late for the development of much of this technology.

422 Written evidence from Kemp Little LLP ([AIC0133](#))

423 *Ibid.*

424 *Ibid.*

425 [Q 32](#) (Professor Chris Reed)

426 *Ibid.*

427 [Q 32](#) (Professor Karen Yeung)

428 *Ibid.*

429 [Q 111](#) (Paul Clarke)

430 [Q 111](#) (Dr Mark Taylor)

431 [Q 111](#) (Dr Joseph Reger)

312. euRobotics highlighted the work of the Committee on Legal Affairs (JURI) in the European Parliament in this area. JURI established a Working Group on legal questions related to the development of robotics and artificial intelligence in the European Union on 20 January 2015. The resulting report, *Civil Law Rules on Robotics*, was published on 27 January 2017.⁴³² The report made recommendations to the European Commission and called for EU-wide rules for robotics and artificial intelligence, in order to fully exploit their economic potential and to guarantee a standard level of safety and security.
313. JURI requested, amongst many recommendations, that draft legislation to clarify liability issues (in particular for driverless cars), and for a mandatory insurance scheme and supplementary fund to compensate victims of accidents involving self-driving cars. The Commission was also asked to consider giving legal status to robots, in order to establish who is liable if they cause damage. On 16 February 2017, the European Parliament adopted JURI's report.⁴³³
314. Our witnesses also raised the issue of legal personality—the term used to establish which entities have legal rights and obligations, and which can do such things as enter into contracts or be sued⁴³⁴—for artificial intelligence.⁴³⁵ A group of academic witnesses said “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”.⁴³⁶
315. Dr Sarah Morley and Dr David Lawrence, both of Newcastle University Law School, said “the decision to award legal status to AI will have many ramifications for legal responsibility and for issues such as legal liability”.⁴³⁷ On the other hand, Dr Morley and Dr Lawrence told us, “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 (companies) or the owners”.⁴³⁸ They also said that there may be issues for criminal liability.⁴³⁹
316. Professor Yeung said that the issue of whether or not algorithms should have legal personality “must be driven by how you envisage the distribution of loss, liability and responsibility more generally”.⁴⁴⁰ Professor Yeung told us that the nature of the compensation system was also important, and that a negligence-based system, relying on a chain of causation—the series of events used to assess liability for damages—would be broken by “the

432 European Parliament, *Report with recommendations to the Commission on Civil Law Rules on Robotics* (27 January 2017): <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//EN> [accessed 12 January 2018]

433 European Parliament, *Resolution on Civil Law Rules on Robotics* (16 February 2017): <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2017-0051+0+DOC+PDF+V0//EN> [accessed 6 March 2018]

434 Further, as defined by the Oxford Legal Dictionary (7th edition, 2014), legal personality is “principally an acknowledgement that an entity is capable of exercising certain rights and being subject to certain duties on its own account under a particular system of law. In municipal systems, the individual human being is the archetypal “person” of the law, but certain entities, such as limited companies or public corporations, are granted a personality distinct from the individuals who create them. Further, they can enter into legal transactions in their own name and on their own account.”

435 Written evidence from Weightmans LLP ([AIC0080](#)).

436 Written evidence from Dr Aysegul Bugra, Dr Matthew Channon, Dr Ozlem Gurses, Dr Antonios Kouroutakis and Dr Valentina Rita Scotti ([AIC0051](#))

437 Written evidence from Dr Sarah Morley and Dr David Lawrence ([AIC0036](#))

438 *Ibid.*

439 *Ibid.*

440 [Q 31](#) (Professor Karen Yeung)

lack of reasonable foresight” offered by an algorithm. Professor Reed was less concerned about this particular issue, and told us “you never apply law to technology; you always apply law to humans and the way they use technology, so there will always be someone who is using the algorithm on whom responsibility can be placed”.⁴⁴¹

317. **In our opinion, it is possible to foresee a scenario where AI systems may malfunction, underperform or otherwise make erroneous decisions which cause harm. In particular, this might happen when an algorithm learns and evolves of its own accord. It was not clear to us, nor to our witnesses, whether new mechanisms for legal liability and redress in such situations are required, or whether existing mechanisms are sufficient.**
318. **Clarity is required. We recommend that the Law Commission consider the adequacy of existing legislation to address the legal liability issues of AI and, where appropriate, recommend to Government appropriate remedies to ensure that the law is clear in this area. At the very least, this work should establish clear principles for accountability and intelligibility. This work should be completed as soon as possible.**

Criminal misuse of artificial intelligence and data

319. There was some concern amongst our witnesses that AI will, and indeed could already be, super-charging conventional cyber-attacks, and facilitating an entirely new scale of cyber-attack.
320. There is some debate within the cybersecurity community as to whether hackers are already using AI for offensive purposes. At the recent Black Hat USA 2017 cybersecurity conference, a poll found that 62% of attendees believed that machine learning was already being deployed by hackers.⁴⁴²
321. The Future of Humanity Institute highlighted the potential use of AI for ‘spear phishing’, a kind of cyber-attack where an email is tailored to a specific individual, organisation or business, usually with the intent to steal data or install malware on a target computer or network.⁴⁴³ Using AI, a normally labour intensive form of cyber-attack could be automated, thereby substantially increasing the number of individuals or organisations that can be targeted.
322. AI systems can also have particular vulnerabilities which do not exist in more conventional systems. The field of ‘adversarial AI’ is a growing area of research, whereby researchers, armed with an understanding of how AI systems work, attempt to fool other AI systems into making incorrect classifications or decisions. In recent years, image recognition systems in particular have been shown to be susceptible to these kinds of attacks. For example, it has been shown that pictures, or even three-dimensional models or signs, can be subtly altered in such a way that they remain indistinguishable from the originals, but fool AI systems into recognising them as completely different objects.⁴⁴⁴

441 Q 31 (Professor Chris Reed)

442 Cylance, ‘Black Hat attendees see AI as double-edged sword’ (1 August 2017): https://www.cylance.com/en_us/blog/black-hat-attendees-see-ai-as-double-edged-sword.html [accessed 23 January 2018]

443 Written evidence from Future of Humanity Institute (AIC0103)

444 Written evidence from Dr Julian Estevez (AIC0021)

323. In written evidence, the Reverend Dr Lyndon Drake gave the following examples:

“ ... 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-machine learning systems too (or indeed with human decision-makers), but with non-machine learning approaches the reasoning involved can be interrogated, recovered, and debugged. This is not possible with many machine learning systems”.⁴⁴⁵

324. Adversarial AI also has implications for AI-enabled approaches to cybersecurity. While we heard from a number of witnesses who argued that AI was already helping to prevent cyber-attacks, some researchers have argued that AI-powered cybersecurity systems might be tricked into allowing malware through firewalls.⁴⁴⁶ NCC Group, a cybersecurity company, told us that the black box nature of most machine learning-based products in use today, which prevents humans understanding much about how data is being processed, means adversaries have:

“ ... 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 machine learning tasks ... We believe that it is inevitable that attackers will start using AI and machine learning 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 machine learning in their attacks, they will”.⁴⁴⁷

325. However, it is not yet clear how serious this problem is likely to be in real-world scenarios. Most examples to date have not been considered ‘robust’—while they may fool an AI system from a particular angle, usually if an image is rotated or zoomed in slightly, the effect is lost. Recent experiments, however, have shown the possibility of creating more robust attacks. Many AI developers have started to consider adversarial hacking, and in some cases are considering possible countermeasures. When we asked Professor Chris Hankin, Director of the Institute for Security Science and Technology at Imperial College, about the implications of this, he informed us that “at the moment certainly, AI is not the only answer we should be thinking about for defending our systems”.⁴⁴⁸

326. During our visit to Cambridge, researchers from the Leverhulme Centre for the Future of Intelligence said that many developments in AI research have many different applications, which can be put to good use, but can equally be abused or misused. They claimed that AI researchers can often be naïve about the possible applications of their research. They suggested that a very small percentage (around 1%) of AI research regarding applications with a high risk of misuse should not be published, on the grounds that the risks outweighed the benefits. As much AI research, even in some cases from large corporations, is published on an open access or open source basis, this would contravene the general preference for openness among many AI researchers.

445 Written evidence from the Reverend Dr Lyndon Drake ([AIC0108](#))

446 Written evidence from NCC Group plc ([AIC0240](#)) and [Q 147](#) (Professor Chris Hankin)

447 Written evidence from NCC Group plc ([AIC0240](#))

448 [Q 147](#) (Professor Chris Hankin)

327. When we put this to Dr Mark Briers, Strategic Programme Director for Defence and Security, Alan Turing Institute, he said that “there is an ethical responsibility on all AI researchers to ensure that their research output does not lend itself to obvious misuse and to provide mitigation, where appropriate”.⁴⁴⁹ However, drawing on the example of 3D printing, where the same technology that can illicitly produce firearms is also producing major medical advances, he believed that “principles and guidelines, as opposed to definitive rules” were more appropriate. Professor Hankin also agreed with this approach, noting precedents in cybersecurity, where many vendors provide ‘bug bounties’ to incentivise the disclosure of security vulnerabilities in computer systems by researchers and other interested parties, so that they can be patched before knowledge of their existence is released into the public domain.⁴⁵⁰
328. **The potential for well-meaning AI research to be used by others to cause harm is significant. AI researchers and developers must be alive to the potential ethical implications of their work. The Centre for Data Ethics and Innovation and the Alan Turing Institute are well placed to advise researchers on the potential implications of their work, and the steps they can take to ensure that such work is not misused. However, we believe additional measures are required.**
329. **We recommend that universities and research councils providing grants and funding to AI researchers must insist that applications for such money demonstrate an awareness of the implications of the research and how it might be misused, and include details of the steps that will be taken to prevent such misuse, before any funding is provided.**
330. Witnesses also told us that the potential misuse of AI should be considered in terms of the data fed into these systems. The subject of adversarial attacks illustrates more widely how misleading data can also harm the integrity of AI systems. In Chapter 3 we considered the issue of how biased datasets can lead an AI system to the wrong conclusions, but systems can also be corrupted on purpose. As 10x Future Technology put it, “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”.⁴⁵¹ NCC Group said: “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”.⁴⁵²
331. There are a number of possible solutions to these issues. NCC Group highlighted recent research on countering adversarial attacks by devising means to detect and reject ‘dangerous data’ before it can reach the classification mechanisms of an AI system.⁴⁵³ However, they believed it was more important to ensure that the data used to train and operate AI systems were not put at risk of interference in the first place, and “to counter such risks, clear processes and mechanisms need to be in place by

449 [Q 146](#) (Dr Mark Briers)

450 [Q 146](#) (Professor Chris Hankin). ‘Bug bounties’ are monetary rewards paid by software companies to individuals who find and disclose vulnerabilities to them.

451 Written evidence from 10x Future Technology ([AIC0024](#))

452 Written evidence from NCC Group plc ([AIC0240](#))

453 *Ibid.*

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".⁴⁵⁴ They suggested that mandatory third-party validation of AI systems should be considered, in order to periodically check their effectiveness, especially in the case of cybersecurity systems which are safeguarding other systems.

332. We note these concerns, and are surprised that the Cabinet Office's recently published Interim Cyber Security Strategy, while making reference to the opportunities for deploying AI in cybersecurity contexts, does not make any mention of the associated risks.⁴⁵⁵ This is particularly important given the current push for more Government data to be opened up for public use, and we are convinced that measures must be put in place to ensure that the integrity and veracity of this data is not corrupted, and that advice is provided to the private sector to ensure their datasets are similarly protected against malicious use.
333. **We recommend that the Cabinet Office's final Cyber Security Science & Technology Strategy take into account the risks as well as the opportunities of using AI in cybersecurity applications, and applications more broadly. In particular, further research should be conducted into methods for protecting public and private datasets against any attempts at data sabotage, and the results of this research should be turned into relevant guidance.**

Autonomous weapons

334. Perhaps the most emotive and high-stakes area of AI development today is its use for military purposes. While we have not explored this area with the thoroughness and depth that only a full inquiry into the subject could provide, there were particular aspects which needed acknowledging, even in brief. The first distinction that was raised by witnesses was between the relatively uncontroversial use of AI for non-violent military applications, such as logistics and strategic planning, and its use in autonomous weapons, or so-called 'killer robots'. The former uses, while representing an area of substantial innovation and growth at the moment, were deemed uncontentious by all of our witnesses and the issues they present appear to be broadly in alignment with the ethical aspects of civilian AI deployment.⁴⁵⁶ As such, we have chosen to focus exclusively on the issue of autonomous weaponry.
335. We quickly discovered that defining the concept of autonomous weaponry with any precision is fraught with difficulty. The term cannot simply be applied to any weapon which makes use of AI; indeed, as one respondent pointed out, "no modern anti-missile system would be possible without the use of AI systems".⁴⁵⁷ Most witnesses used the term to describe weapons which autonomously or semi-autonomously target or deploy violent force, but within this there are many shades of grey, and our witnesses disagreed with one another when it came to describing autonomous weapons in use

454 *Ibid.*

455 Cabinet Office, *Interim cyber security science & technology strategy: Future-proofing cyber security* (December 2017), pp 8–9: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/663181/Embargoed_National_Cyber_Science_and_Technology_Strategy_FINAL.pdf [accessed 30 January 2018]

456 Q 154 (Mike Stone, Professor Noel Sharkey, Major Kitty McKendrick, Dr Alvin Wilby)

457 Written evidence from the Reverend Dr Lyndon Drake (AIC0108)

or development. For example, Dr Alvin Wilby, Vice-President of Research, Technical and Innovation, Thales, suggested that the Israeli Harpy drone, which is “capable of loitering over an area and deciding which target to go for”, and has already been deployed in armed conflict, would count as an autonomous weapon.⁴⁵⁸ But Professor Noel Sharkey, Professor of robotics and artificial intelligence, University of Sheffield, disputed this definition, arguing that the Harpy drone was a relatively simple computational system, and demonstrated the extent to which the “the term AI is running out of control” within the arms industry.⁴⁵⁹ Professor Sharkey also highlighted how arms manufacturers had a tendency to play up the sophistication and autonomy of their products in marketing, and downplay them when scrutinised by international bodies such as the United Nations (UN).

336. It was generally agreed that the level of human control or oversight over these weapons was at the heart of the issue. While it is now common to simply refer to there being a ‘human in the loop’ with many semi-autonomous weapons in use or active development, it emerged that this could mean many things. Professor Sharkey outlined a number of different levels of autonomy.⁴⁶⁰ This ranged from ‘fire-and-forget’ missiles, such as the Brimstone missile used by UK armed forces, which have a pre-designated target but can change course to seek this out; through to more autonomous systems with a brief ‘veto period’, such as the US Patriot missile system, in which a human can override the automated decision; and finally fully autonomous weapons, which seek out their own targets without human designation.

Box 10: UK Government definitions of automated and autonomous systems

The Ministry of Defence (MoD) most recently defined autonomous weapons in official guidance on unmanned aircraft systems in September 2017, and has made a relatively unusual distinction between automated and autonomous systems.

Automated system

In the unmanned aircraft context, 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.

Autonomous system

An autonomous system is 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, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be.

Source: Ministry of Defence, *Unmanned aircraft systems (12 September 2017)*: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf [accessed 7 February 2018]

458 [Q 154](#) (Dr Alvin Wilby)

459 [Q 157](#) (Professor Noel Sharkey)

460 [Q 155](#) (Professor Noel Sharkey)

337. The distinctions, whilst technical, take on a greater significance given the current moves to place restrictions on autonomous weapons under international law. At a meeting of experts, convened by the UN, in April 2016, 94 countries recommended beginning formal discussions about lethal autonomous weapons systems. The talks are to consider whether these systems should be restricted under the Convention on Certain Conventional Weapons, a disarmament treaty that has regulated or banned several other types of weapons, including incendiary weapons and blinding lasers. In November 2017, 86 countries participated in a meeting of the UN's Convention on Certain Conventional Weapons Group of Governmental Experts. 22 countries now support a prohibition on fully autonomous weapons, including, most recently, Brazil, Uganda and Iraq.⁴⁶¹
338. In September 2017, the MoD issued updated guidance stating that “the UK does not possess fully autonomous weapon systems and has no intention of developing them. Such systems are not yet in existence and are not likely to be for many years, if at all”.⁴⁶² It is important to note that the UK distinguishes between ‘autonomous’ and ‘automated’ military systems (see Box 10).
339. The Government has also opposed the proposed international ban on the development and use of autonomous weapons. The Government argues that existing international human rights law is adequate, and that the UN Convention on Certain Conventional Weapons currently allows for adequate scrutiny of automated and autonomous weapons under its mechanisms for legal weapons review.⁴⁶³
340. Professor Sharkey argued that requiring an autonomous weapon system to be “aware and show intention”, as stated in MoD guidance, was to set the bar so high that it was effectively meaningless.⁴⁶⁴ He also told us that it was “out of step” with the rest of the world, a point seemingly acknowledged by the MoD in their guidance, which states that “other countries and industry often have very different definitions or use the terms [autonomous and automated] interchangeably”.⁴⁶⁵
341. In practice, this lack of semantic clarity could lead the UK towards an ill-considered drift into increasingly autonomous weaponry. Professor Sharkey noted that BAE Systems has described its Taranis unmanned vehicle as ‘autonomous’, and that this capacity has been “widely tested in Australia”.⁴⁶⁶ On the other hand, Major Kitty McKendrick, speaking in her capacity as a visiting fellow at Chatham House, argued that she would not consider systems that have been told to “look for certain features [and identify targets] on that basis” as “genuinely autonomous” as they are acting “in accordance with a predictable program”.⁴⁶⁷

461 ‘Support grows for new international law on killer robots’, *Campaign to stop killer robots* (17 November 2017): <https://www.stopkillerrobots.org/2017/11/gge/> [accessed 1 February 2018]

462 Ministry of Defence, *Unmanned Aircraft Systems* (September 2017), p 14: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf [accessed 18 January 2018]

463 *Unmanned Aircraft Systems*, p 43

464 Q 155 (Professor Noel Sharkey)

465 *Unmanned Aircraft Systems*, p 20

466 Q 156 (Professor Noel Sharkey)

467 Q 157 (Major Kitty McKendrick)

Box 11: Definitions of lethal autonomous weapons systems used by other countries

The following are definitions of lethal autonomous weapons systems (LAWS) used by other countries.

Austria

Autonomous weapons systems (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.

France

LAWS should be understood as implying a total absence of human supervision, meaning there is absolutely no link (communication or control) with the military chain of command. The delivery platform of a LAWS would be capable of moving, adapting to its land, marine or aerial environments and targeting and firing a lethal effector (bullet, missile, bomb, etc.) without any kind of human intervention or validation.

The Holy See

An autonomous weapon system is a weapon system capable of identifying, selecting and triggering action on a target without human supervision.

Italy

LAWS 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.

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.

Norway

Weapons that would search for, identify and attack targets, including human beings, using lethal force without any human operator intervening.

Switzerland

AWS are weapons systems that are capable of carrying out tasks governed by international humanitarian law, in partial or full replacement of a human in the use of force, notably in the targeting cycle.

USA

A weapon system that, once activated, can select and engage targets without further intervention by a human operator.

Source: Written evidence from Professor Noel Sharkey ([AIC0248](#))

342. While the definitions in Box 11, which mostly represent NATO-member countries, vary in their wording, none would appear to set the bar as high as the UK. All of these definitions focus on the level of human involvement in supervision and target setting, and do not require “understanding higher-level intent and direction”, which could be taken to mean at least some level of sentience.

343. When we asked Matt Hancock MP about the UK's definition, he said:

“There is not an internationally agreed definition of lethal autonomous weapons systems. We think that the existing provisions of international humanitarian law are sufficient to regulate the use of weapons systems that might be developed in the future. Of course, having a strong system and developing it internationally within the UN Convention on Certain Conventional Weapons is the right way to discuss the issue. Progress was made in Geneva by the group of government experts just last month. It is an important area that we have to get right”.⁴⁶⁸

344. We were encouraged by the Minister's willingness to enter into this debate, and consider the need for a change in Government policy in this area. Regardless of the merits or otherwise of an international ban, as Mike Stone, former Chief Digital and Information Officer, Ministry of Defence, emphasised there is a need for a “very clear lexicon” in this area which does not necessarily apply in most civilian domains.⁴⁶⁹

345. **Without agreed definitions we could easily find ourselves stumbling through a semantic haze into dangerous territory. The Government's definition of an autonomous system used by the military as one where it “is capable of understanding higher-level intent and direction” is clearly out of step with the definitions used by most other governments. This position limits both the extent to which the UK can meaningfully participate in international debates on autonomous weapons and its ability to take an active role as a moral and ethical leader on the global stage in this area. Fundamentally, it also hamstring attempts to arrive at an internationally agreed definition.**

346. **We recommend that the UK's definition of autonomous weapons should be realigned to be the same, or similar, as that used by the rest of the world. To produce this definition the Government should convene a panel of military and AI experts to agree a revised form of words. This should be done within eight months of the publication of this report.**

468 [Q 199](#) (Matt Hancock MP)

469 [Q 155](#) (Mike Stone)

CHAPTER 9: SHAPING ARTIFICIAL INTELLIGENCE

347. Over the course of our inquiry the Government has made a series of announcements regarding artificial intelligence, mostly in the Industrial Strategy and in response to the Hall-Pesenti Review. These policies are nascent, but welcome. They represent the Government’s commitment to AI, with a focus on the AI development sector in the UK. We believe these policies are a good base upon which to build, and for the Government to show strong domestic leadership on AI. Our recommendations in this chapter focus on the action the Government can take to maximise the potential of AI for the UK, and to minimise its risks.⁴⁷⁰

Leading at home

348. Much of the recent policy focus by the Government has related to the announcement of a series of new AI-related bodies.

The AI Council and the Government Office for AI

349. The Hall-Pesenti Review recommended that the “Government should work with industry and experts to establish a UK AI Council to help coordinate and grow AI in the UK.” The recommendation was based on the perceived need to facilitate engagement between industry, academia, Government and the public, as “AI in the UK will need to build trust and confidence in AI-enabled complex systems”.⁴⁷¹

350. In the Industrial Strategy, the Government announced that they were taking forward this recommendation, and “working with industry to establish an industry-led AI Council that can take a leadership role across sectors”.⁴⁷² It was announced that the Council would be supported by a new Government Office for AI. The Industrial Strategy stated that both these bodies would:

- champion research and innovation;
- stimulate demand and accelerate uptake across all sectors of the economy;
- increase awareness of the advantages of advanced data analytic technologies; and
- promote greater diversity in the AI workforce.⁴⁷³

351. It was not clear from the announcements how these new bodies would be constituted, or how they might function. We asked Matt Hancock MP who would be represented on the AI Council. He told us “there has to be small and medium-sized business representation but also users and developers”.⁴⁷⁴ Dr Pesenti told us the Council “should be a mix of industry, academia and Government”.⁴⁷⁵ Dr Pesenti told us that the Council should not be too big, but that it should have “one or two representatives” of large companies and

470 Appendix 9 to this report shows which of our recommendations is relevant to which newly established AI-related body.

471 *Growing the artificial intelligence industry in the UK*, p 5

472 *Industrial Strategy: Building a Britain fit for the future*, p 39

473 *Ibid.*

474 [Q 193](#) (Matt Hancock MP)

475 [Q 206](#) (Dr Jérôme Pesenti)

of start-ups in AI.⁴⁷⁶ Dr Pesenti also said that membership should be “UK-centric” although he “would put international companies on it. You could have people with an interest in the UK who are part of these companies”.⁴⁷⁷ Dr Pesenti also told us that he envisaged the Council reporting annually on the progress made by the UK against agreed metrics, and playing a role in ensuring that the aims of policies are being delivered.⁴⁷⁸

352. Matt Hancock MP told us the Government Office for AI would be a “joint unit between BEIS and DCMS to ensure that we are joined up at the central Government level”.⁴⁷⁹ The Minister said “we think it will be resourced by civil servants reporting directly to Ministers ... It is the team that will manage this policy development and architecture”.⁴⁸⁰

Centre for Data Ethics and Innovation

353. As we have previously discussed, the Industrial Strategy also announced a new Centre for Data Ethics and Innovation. This would be a “world-first advisory body” which would review the current “governance landscape” and advise the Government on “ethical, safe and innovative uses of data, including AI”.⁴⁸¹ The Centre would engage with industry to establish data trusts, and there would be wide consultation as to the remit of the Centre in due course.⁴⁸² The Prime Minister reaffirmed these ambitions in her speech to the World Economic Forum on 25 January 2018.⁴⁸³
354. Matt Hancock MP told us that the Centre “will not be a regulatory body, but it will provide the leadership that will shape how artificial intelligence is used”.⁴⁸⁴ The Minister said the Government wanted “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”.⁴⁸⁵ The Minister cited the Human Fertilisation and Embryology Authority as an example of how this can be an effective approach (see Box 12), and said “it is incredibly important to ensure that society moves at the same pace as the technology, because this technology moves very fast”.⁴⁸⁶

476 *Ibid.*

477 *Ibid.*

478 *Ibid.*

479 [Q 191](#) (Matt Hancock MP)

480 *Ibid.*

481 *Industrial Strategy: Building a Britain fit for the future*, p 40

482 *Ibid.*

483 Prime Minister Theresa May, Speech at the World Economic Forum in Davos, Switzerland, 25 January 2018: <https://www.gov.uk/government/speeches/pms-speech-at-davos-2018-25-january> [accessed 1 March 2018]

484 [Q 191](#) (Matt Hancock MP)

485 *Ibid.*

486 *Ibid.*

Box 12: The Human Fertilisation and Embryology Authority

The Human Fertilisation and Embryology Authority (HFEA) was established as a non-departmental public body by the Human Fertilisation and Embryology Act 1990, and came into being on 1 August 1991. The Human Fertilisation and Embryology Act 2008 updated the role of the HFEA. The 1990 Act was the result of a report into the issues surrounding in vitro fertilisation (IVF) by a committee chaired by Baroness Warnock.

Baroness Harding of Winscombe cited the work of the Warnock Committee as an example of where “we have had the moral and ethical debate as technology was developing and the possibilities the technology offered began to emerge”, and that the work of the Committee “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”.⁴⁸⁷

Source: Human Fertilisation & Embryology Authority, ‘About us’: <https://www.hfea.gov.uk/about-us/> [accessed 8 February 2018]

355. The remit of the Centre the Minister outlined to us was extensive, and included:

- working with experts to develop an ethical framework (and publicising that work);
- being a public advocate of the benefit of the technology;
- recommending changes to policy where appropriate;
- the promotion of standards around the use of data; and
- developing data trusts (trusted mechanisms to make it easier for organisations to understand how to use and share data for AI safely and securely).⁴⁸⁸

356. The data trusts which the Centre could be charged with the development of are another recommendation of the Hall-Pesenti Review, which said:

“To facilitate the sharing of data between organisations holding data and organisations looking to use data to develop AI, Government and industry should deliver a programme to develop data trusts—proven and trusted frameworks and agreements—to ensure exchanges are secure and mutually beneficial”.⁴⁸⁹

357. The Review envisaged the data trusts as being “a set of relationships underpinned by a repeatable framework, compliant with parties’ obligations, to share data in a fair, safe and equitable way”.⁴⁹⁰ Dr Pesenti told us that “one size definitely does not fit all when you share data” and the Review’s concept of data trusts was to establish a trusted way to facilitate the “sharing of data among multiple parties.” He told us that phase one of establishing the trusts would entail trialling agreements and working out what could work as a template. Phase two would involve looking to reach a point where: “You, as a person, do not give your data to an organisation. Even when you go to a

487 Written evidence from Baroness Harding of Winscombe (AIC0072)

488 Q 196 (Matt Hancock MP)

489 *Growing the artificial intelligence industry in the UK*, p 4

490 *Ibid.*

website, when they collect that data, they do not put it in their organisation. They do not own that data, but they put it in a trust, where it is very visible and clear how that data will be used”.⁴⁹¹

358. This is an extensive, and potentially challenging, remit for any organisation, let alone a newly established one. The Minister himself warned us that he did not want to “encumber it with so much at the start that we are trying to boil the ocean”.⁴⁹² In January 2018, the Government advertised for a Chair for an interim Centre, the role of which would be to “start work on key issues straight away” with its findings to be “used to inform the final design and work programme of the permanent Centre”.⁴⁹³ The advertisement also noted that the Government intends to establish the Centre “on a statutory footing in due course”.⁴⁹⁴ The citing of HFEA as an example, and the decision to place the Centre on a statutory basis, suggest that the Centre will not be dissimilar to a regulator. Care must be taken to not do this inadvertently. The plans for the Centre at this stage can, of course, be considered as a possible blueprint for a regulator if, and when, one is necessary.

A National Institute for AI

359. At the same time as the announcement of the creation of the Centre for Data Ethics and Innovation, the AI Council, and the Government Office of AI, it was announced that the Alan Turing Institute would become the national research centre for AI, and would be supporting Turing Fellowships, doctoral studentships linked to the Institute and focused on AI-related issues. This was another of the Hall-Pesenti Review’s recommendations.
360. The Alan Turing Institute is a national centre for the study of data science established in 2015. In March 2014 the Government announced they were dedicating £42 million to establish the Alan Turing Institute, and the institute also receives funding from its founding universities, donations and grants.
361. Matt Hancock MP explained to us the need for the national centre for AI alongside the other AI bodies being established. The Minister said that the Alan Turing Institute was “essentially the champion of artificial intelligence research” and that as such did “not want the university-led champion of AI research also to be the body that does the thinking on the ethics and framework, because ... we want the Alan Turing Institute to be able to take on industrial sponsorship ... and work directly for corporates in developing AI”.⁴⁹⁵
362. We asked Dr Pesenti why the Alan Turing Institute was recommended as the national centre. Dr Pesenti told us it was because of the Institute’s name: “Turing is one of the most recognised names in AI and it is a great legacy to what has been done here. You cannot just go around it. You are not going to create a new institute that is Turing II”.⁴⁹⁶ Dr Pesenti also told us that “there is scepticism in the industry that the institute is where it should be, in terms of efficacy and delivery, so it is really important for an institute to step up”.⁴⁹⁷

491 [Q 207](#) (Dr Jérôme Pesenti)

492 [Q 196](#) (Matt Hancock MP)

493 CentreforPublicAppointments, ‘InterimcentreforDataEthicsandInnovation: Chair’ (25January2018): <https://publicappointments.cabinetoffice.gov.uk/appointment/interim-centre-data-ethics-innovation-chair/> [accessed 1 February 2018]

494 *Ibid.*

495 [Q 191](#) (Matt Hancock MP)

496 [Q 208](#) (Dr Jérôme Pesenti)

497 *Ibid.*

We asked Lord Henley whether the Alan Turing Institute had the capacity to act in the role envisaged for it. Lord Henley said “where we do not really know what will happen, it is best to let a thousand things bloom so that the Government, as long as they remain nimble, can respond in the appropriate way”.⁴⁹⁸ Dr Barber, who is a Turing Fellow at the Alan Turing Institute, told us that the Institute had the appetite to focus more on AI, but that the “level of commitment would require financial support and it cannot easily be done with the current resources at the Institute”.⁴⁹⁹

363. The UK is not the only country to have taken steps to establish national centres for artificial intelligence research. In 2017, the Canadian government’s ‘Pan-Canadian Artificial Intelligence Strategy’ looked to establish “nodes of scientific excellence in Canada’s three major centres for artificial intelligence in Edmonton, Montreal and Toronto-Waterloo”.⁵⁰⁰ The most prominent of these currently is the Vector Institute. It will be hosted by the University of Toronto, and funded by the governments of Ontario and Canada (C\$90 million), and the private sector (C\$80 million). Geoff Hinton, an early pioneer of deep learning, will be the Vector Institute’s chief scientific advisor. There is much potential for the Alan Turing Institute to learn from the experiences of other national centres for AI, and we urge the Institute to develop a relationship with the new centres being established in Canada, and elsewhere, in order to meet the challenge with which it has been presented. The work of the German Research Center for AI (DFKI), to whom we spoke, offers another source of potential advice, as a much longer standing centre (having been founded in 1988).
364. We welcome the enthusiasm, and speed, with which the Government has met the emergence of artificial intelligence. However, we agree with Nicola Perrin who said to us that “a proliferation of different bodies has been proposed” and that “there needs to be much greater clarity in the governance landscape”.⁵⁰¹ Nicola Perrin asked us to “give clarity over what is needed rather than suggesting yet another new body, it would be very helpful”.⁵⁰² This was before the announcements of the bodies above were made in the 2017 Budget and Industrial Strategy. We have given consideration to this when making our own recommendations (see Appendix 9).
365. Many of our witnesses called for an AI-specific national strategy. Professor Hall, speaking to us prior to the publication of her review with Dr Pesenti, said “I hope when you see the review you will think there are the beginnings of that strategy there”.⁵⁰³ Dr Taylor wanted a “comprehensive strategy around how the country is going to benefit from its exploitation”.⁵⁰⁴ Sage said “without a clear AI strategy for social good the huge potential benefits of AI will not be felt across society at large”.⁵⁰⁵ It is clear to us that with the Government being so actively engaged with AI, and the number of institutes that could now possibly be involved in shaping and developing AI policy, that a clear framework is required.

498 [Q 191](#) (Lord Henley)

499 [Q 42](#) (Dr David Barber)

500 CIFAR, ‘Pan-Canadian Artificial Intelligence Strategy Overview’ (30 March 2017): <https://www.cifar.ca/assets/pan-canadian-artificial-intelligence-strategy-overview/> [accessed 29 January 2018]

501 [Q 127](#) (Nicola Perrin)

502 *Ibid.*

503 [Q 8](#) (Professor Dame Wendy Hall)

504 [Q 115](#) (Dr Mark Taylor)

505 Written evidence from Sage ([AIC0159](#))

366. **Artificial intelligence’s potential is an opportunity the Government is embracing. The Government’s recent enthusiasm and responsiveness to artificial intelligence in the UK is to be welcomed. We have proposed a number of recommendations for strengthening recent policy announcements, based on the extensive evidence we have received as a Committee. We encourage the Government to continue to be proactive in developing policies to harness the potential of AI and mitigate the risks. We do, however, urge the Government to ensure that its approach is focused and that it provides strategic leadership—there must be a clear roadmap for success. Policies must be produced in concert with one another, and with existing policy. Industry and the public must be better informed about the announcements, and sufficient detail provided from the outset.**
367. **The pace at which this technology will grow is unpredictable, and the policy initiatives have been many. To avoid policy being too reactive, and to prevent the new institutions from overlapping and conflicting with one another, we recommend that the Government Office for AI develop a national policy framework for AI, to be in lockstep with the Industrial Strategy, and to be overseen by the AI Council. Such a framework should include policies related to the recommendations of this report, and be accompanied, where appropriate, by a long-term commitment to such policies in order to realise the benefits. It must also be clear within Government who is responsible around the Cabinet table for the direction and ownership of this framework and the AI-related policies which fall within it.**
368. **The roles and remits of the new institutions must be clear, if they are to be a success. The public and the technology sector in the UK must know who to turn to for authoritative advice when it comes to the development and use of artificial intelligence. To ensure public confidence, it must also be clear who to turn to if there are any complaints about how AI has been used, above and beyond the matters relating to data use (which falls within the Information Commissioner’s remit).**
369. **We recommend that the Government Office for AI should act as the co-ordinator of the work between the Centre for Data Ethics and Innovation, the GovTech Catalyst team and the national research centre for Artificial Intelligence Research (the Alan Turing Institute), as well as the AI Council it is being established to support. It must also take heed of the work of the more established bodies which have done work in this area, such as the Information Commissioner’s Office and the Competition and Markets Authority. The work programmes of all the new AI-specific institutions should be subject to agreement with one another, on a quarterly basis, and should take into account the work taking place across Government in this area, as well as the recommendations from Parliament, regulators, and the work of the devolved assemblies and governments. The UK has a thriving AI ecosystem, and the Government Office for AI should seek to inform its work programme through wide public consultation as it develops Government policy with regard to artificial intelligence. The programme should be publicly available for scrutiny.**

370. **We welcome the new focus for the Alan Turing Institute as the national research centre for artificial intelligence. We want it to be able to fulfil this role, and believe it has the potential to do so. As such, the new focus must not simply be a matter of rebranding. The successful institutes in Canada and Germany, such as the Vector Institute and the German Research Center for Artificial Intelligence, offer valuable lessons as to how a national research centre should be operated.**
371. **The Government must ensure that the Alan Turing Institute’s funding and structure is sufficient for it to meet its new expanded remit as the UK’s national research centre for AI. In particular, the Institute’s current secondment-based staffing model should be assessed to ensure its suitability, and steps taken to staff the Institute appropriately to meet the demands now placed upon it.**

Regulation and regulators

372. The role the Centre for Data Ethics and Innovation is to play in reviewing the current governance landscape is a challenging one. Many of our witnesses commented on the desirability of regulation (or not), including whether a blanket AI-specific regulation is required. Our witnesses also commented on the need for a specific regulator was required, or whether the existing regulatory landscape was sufficient.
373. Witnesses fell into three broad camps: those who considered existing laws could do the job; those who thought that action was needed immediately; and those who proposed a more cautious and staged approach to regulation. Those who said that no new AI-specific regulation was required did so on the basis that existing laws and regulations could adequately regulate the development and use of AI.⁵⁰⁶ TechUK, who represent over 950 companies, told us that “the concerns regarding the use of AI technologies ... are focused around how data is being used in these systems” and that it was “important to remember that the current data protection legal framework is already sufficient”, and the GDPR would further strengthen that framework.⁵⁰⁷ Further, techUK advocated a cautious approach to other areas where regulation might be required:

“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”.⁵⁰⁸

374. The Law Society of England and Wales told us “that there is no obvious reason why the growth of AI and the use of data would require further legislation or regulation”.⁵⁰⁹ They added: “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

506 See written evidence from The Alan Turing Institute ([AIC0139](#)); Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams ([AIC0029](#)); Electronic Frontier Foundation ([AIC0199](#)); techUK ([AIC0203](#)); Arm ([AIC0083](#)); Deep Science Ventures ([AIC0167](#)) and Professor Chris Reed ([AIC00055](#))

507 Written evidence from techUK ([AIC0203](#))

508 *Ibid.*

509 Written evidence from the Law Society of England and Wales ([AIC0152](#))

are any genuine regulatory gaps”.⁵¹⁰ Professor Robert Fisher et al said “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”.⁵¹¹ Professor Bradley Love, a Turing Fellow at the Alan Turing Institute, agreed with this, and said “existing laws and regulations may adequately cover AI” and that “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”.⁵¹²

375. Many others agreed, and some referred to this as a ‘technology-neutral’ approach: “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”.⁵¹³ The Online Dating Association said “the pace of change can make the regulation of technologies, such as AI, very difficult. However, the outputs can be more clearly covered”.⁵¹⁴ They cited the UK’s strong data protection regime, and that “consumer law provides protections around contracts, unfair behaviours, advertising, payments and other critical areas” as examples of an outcomes-focused approach.⁵¹⁵
376. Those arguing for a more cautious approach told us that poorly thought through regulation could have unintended consequences, including the stifling of development, innovation and competitiveness. Professor Love told us that there was a risk that “AI specific regulation could reduce innovation and competitiveness for UK industry” as the competitive advantage gained by using artificial intelligence might be outweighed by regulatory burdens.⁵¹⁶ Dr Reger said:
- “Governments in general—the UK Government might be an exception, and I hope they are—like to regulate. AI technology does not need regulation because it is a competitive race and the faster the United Kingdom progresses in that race, the better it is for the country”.⁵¹⁷
377. Kemp Little LLP said that “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” and that lessons from regulating previous technologies suggested that a “strict and detailed legal requirements approach is unhelpful”.⁵¹⁸ Many other witnesses expressed similar concerns at the possible detrimental effect of premature regulation.⁵¹⁹
378. Baker McKenzie, an international law firm, recommended a “proactive, principles-led intervention, based on a sound understanding of the issues and technology, careful consideration and planning” rather than reactive

510 *Ibid.*

511 Written evidence from Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams ([AIC0029](#))

512 Written evidence from The Alan Turing Institute ([AIC0139](#))

513 Written evidence from SCAMPI Research Consortium, City, University of London ([AIC0060](#))

514 Written evidence from Online Dating Association ([AIC0110](#))

515 *Ibid.*

516 Written evidence from The Alan Turing Institute ([AIC0139](#))

517 [Q 108](#) (Dr Joseph Reger)

518 Written evidence from Kemp Little LLP ([AIC0133](#))

519 See written evidence from Electronic Frontier Foundation ([AIC0199](#)); Professor Chris Reed ([AIC0055](#)) and Professor Robert Fisher, Professor Alan Bundy, Professor Simon King, Professor David Robertson, Dr Michael Rovatsos, Professor Austin Tate and Professor Chris Williams ([AIC0029](#))

regulation, put in place after something goes wrong.⁵²⁰ They recommended that “the right regulatory approach ... is staged and considered” and the Government should “facilitate ethical (as opposed to legal) frameworks for the development of AI technologies” to support self-regulation in industry.⁵²¹

379. There were those who argued for immediate action and regulation, mostly in order to avoid unintended consequences. Dr Morley and Dr Lawrence said “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”.⁵²² The Foundation for Responsible Robotics said “we need to act now to prevent the perpetuation of injustice” and that “there are no guarantees of unbiased performance” for algorithms presently.⁵²³ Bristows LLP argued that it could help with the adoption of the technology, and said:

“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”.⁵²⁴

380. Few witnesses, however, gave any clear sense to us as to what specific regulation should be considered.
381. We heard a number of persuasive arguments against a specific AI-regulation. Professor Reed told us that he “would not have any one-size-fits-all answer” to the question about AI regulation and that it is “inappropriate and impossible to attempt to produce a regulatory regime which applies to *all* AIs”.⁵²⁵ Dr Jerry Fishenden, Visiting Professor at the University of Surrey, said “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”.⁵²⁶ Professor Wooldridge said “AI-specific legislation is not the right way to go. I would look at our existing data protection legislation and ask what AI adds into this mix that we need to start thinking about”.⁵²⁷
382. We were told by the BBC that “the rapid development of AI requires lawmakers and regulators to keep any AI framework under review and up-to-date”.⁵²⁸ Some of our witnesses called for a new, specific artificial intelligence regulator to do this. The Observatory for Responsible Research and Innovation in ICT (ORBIT) said “it will be difficult to regulate AI via straightforward legislation, given the volatile and dynamic nature of this technology” and that it “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”.⁵²⁹ Big Brother

520 Written evidence from Baker McKenzie ([AIC0111](#))

521 *Ibid.*

522 Written evidence from Dr Sarah Morley and Dr David Lawrence ([AIC0036](#))

523 Written evidence from the Foundation for Responsible Robotics ([AIC0188](#))

524 Written evidence from Bristows LLP ([AIC0097](#))

525 [Q 31](#) (Professor Chris Reed) and written evidence from Professor Chris Reed ([AIC0055](#))

526 Written evidence from Dr Jerry Fishenden ([AIC0028](#))

527 [Q 5](#) (Professor Michael Wooldridge)

528 Written evidence from the BBC ([AIC0204](#))

529 Written evidence from ORBIT The Observatory for Responsible Research and Innovation in ICT ([AIC0109](#))

Watch agreed, and called for “independent oversight of AI in the form of a regulatory or supervisory body to provide legal and technical scrutiny of AI technology and algorithms”.⁵³⁰

383. Other witnesses disagreed, arguing that existing regulators were enough. Javier Ruiz Diaz, Policy Director, Open Rights Group, said “a new regulator may end up overlapping with many other regulators”, and that instead “we need to get many regulators to be AI informed and to be able to incorporate AI into their work”.⁵³¹ Olivier Thereaux said “it feels premature to have a regulator for AI. It is probably more useful right now to recognise that AI is going to be used across many sectors, many of which already have regulators”.⁵³² Andrew de Rozairo, Kriti Sharma and James Luke, Chief Technology Officer for the Public Sector, and Master Inventor, IBM, all agreed no new regulator was required, as AI was so intertwined into other business practices that other regulators would suffice.⁵³³ The Government said “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”.⁵³⁴
384. Of existing regulators, AI presents the most pressing issues to the ICO given the current importance of data to machine learning techniques used in AI, and the Data Protection Bill’s proposed changes to the UK’s data protection regime. The ICO upholds information rights in the public interest, promoting openness by public bodies and data privacy for individuals. Elizabeth Denham told us that the GDPR, and the resulting Data Protection Bill:
- “ ... gives us a huge step forward in requiring companies and public bodies to think and to focus on what they are doing with machines, machine learning and artificial intelligence, and to consider the rights of individuals, document that and stand ready to account for the decisions that they have made. The Information Commissioner has the ability to look at those decisions. Individuals have the right to challenge those decisions. We have taken a couple of giant steps forward”.⁵³⁵
385. There is no doubt in our mind that as the development and use of artificial intelligence grows in the UK the ICO will have a pivotal role to play in the regulation of the data underpinning such growth. The Information Commissioner said the ICO’s involvement in the GDPR and changes to data protection law, as well as its involvement in giving (and developing) advice, was feeling “a little like changing a tyre on a moving car”.⁵³⁶ It is essential that the ICO (and other regulators) have the capacity, and support, to fulfil their roles sufficiently. This was recognised by Matt Hancock, who told us that “the ICO is an incredibly important part of getting the new Data Protection Act into place and supporting companies through the changes, and we need to make sure that the Information Commissioner has all the support that she needs to do that”.⁵³⁷ In the Autumn Budget 2017, the Government announced the establishment of the Regulators’ Pioneer Fund

530 Written evidence from Big Brother Watch ([AIC0154](#))

531 [Q 69](#) (Javier Ruiz Diaz)

532 [Q 69](#) (Olivier Thereaux)

533 [Q 83](#) (Andrew de Rozairo, Kriti Sharma, James Luke)

534 Written evidence from HM Government ([AIC0229](#))

535 [Q 60](#) (Elizabeth Denham)

536 [Q 59](#) (Elizabeth Denham)

537 [Q 200](#) (Matt Hancock MP)

“to help unlock the potential of emerging technologies”, which would have £10 million to assist regulators develop innovative approaches for getting products and services to market.⁵³⁸

386. **Blanket AI-specific regulation, at this stage, would be inappropriate. We believe that existing sector-specific regulators are best placed to consider the impact on their sectors of any subsequent regulation which may be needed. We welcome that the Data Protection Bill and GDPR appear to address many of the concerns of our witnesses regarding the handling of personal data, which is key to the development of AI. The Government Office for AI, with the Centre for Data Ethics and Innovation, needs to identify the gaps, if any, where existing regulation may not be adequate. The Government Office for AI must also ensure that the existing regulators’ expertise is utilised in informing any potential regulation that may be required in the future and we welcome the introduction of the Regulator’s Pioneer Fund.**
387. **The additional burden this could place on existing regulators may be substantial. We recommend that the National Audit Office’s advice is sought to ensure that existing regulators, in particular the Information Commissioner’s Office, are adequately and sustainably resourced.**

Assessing policy outcomes

388. Dr Pesenti was clear that the Hall-Pesenti Review was intended to be a first step, and that there were clear metrics in the recommendations of the Review for their implementation to be monitored.⁵³⁹ Dr Andrew Blick said that AI could raise serious questions over the accountability of Ministers and the Government more generally, and require new constitutional models and adjustments to ensure that decisions are properly scrutinised. He suggested that Parliament might consider establishing a committee for artificial intelligence oversight, akin to the Commons Public Accounts Committee. This 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”.⁵⁴⁰ Whilst we do not intend to recommend the establishment of a permanent Parliamentary committee, we do agree with the sentiment that artificial intelligence policy must be scrutinised and the Ministers with responsibility held accountable.
389. One of the central lessons we learnt from historic Government policy on artificial intelligence in the United Kingdom was that a lack of clear short and long-term objective setting for policies in this field can lead to the potential benefits not being realised. Furthermore, a lack of evaluation of these objectives and assessment as the technology grows and develops could prevent the policies from reacting to the uncertain nature of AI.
390. We have made it clear in this report that the growth in the development and use of artificial intelligence offers a significant opportunity for the United Kingdom. There are benefits for society as a whole, the chance to be a

538 *Autumn Budget 2017*, p 49

539 [Q 204](#) (Dr Jérôme Pesenti)

540 Written evidence from Dr Andrew Blick ([AIC0064](#))

world-leader in the development of AI, and potential to shape this emerging technology so that the possible risks are avoided. However, this opportunity will be missed if the Government's commitment to its policies are not sincere.

391. **It is essential that policies towards artificial intelligence are suitable for the rapidly changing environment which they are meant to support. For the UK to be able to realise the benefits of AI, the Government's policies, underpinned by a co-ordinated approach, must be closely monitored and react to feedback from academia and industry where appropriate. Policies should be benchmarked and tracked against appropriate international comparators. The Government Office for AI has a clear role to play here, and we recommend that progress against the recommendations of this report, the Government's AI-specific policies within the Industrial Strategy and other related polices, be reported on an annual basis to Parliament.**

A vision for Britain in an AI world

392. Throughout this report we have discussed the relative strengths and weaknesses of AI development in the UK, but questions still remain regarding Britain's distinctive role in the wider world of AI. The Government has stated in its recent Industrial Strategy White Paper that it intends for the UK to be "at the forefront of the AI and data revolution".⁵⁴¹ What this means in practice is open to interpretation.
393. Some of our respondents appeared to take this at face value, and made comparisons with the United States and China, especially in terms of funding. For example, Nvidia drew attention to the large investments in AI being made in these countries, including the \$5 billion investment announced by the Tianjin state government in China, and the estimated \$20–30 billion investments in AI research from Baidu and Google.⁵⁴² Balderton Capital emphasised the "many billions of funding" being invested in AI and robotics in China and the US, and argued that the UK Government needed to invest more in academic research to ensure that the UK "remains a global leader in the field".⁵⁴³ Microsoft also highlighted the disparities in computer science education, noting that "in a year when China and India each produced 300,000 computer science graduates, the UK produced just 7,000".⁵⁴⁴ Ocado commented favourably on China's relative lack of regulation, observing that "less legislation around the application of technology is fuelling faster experimentation and innovation, including when it comes to the use of data and AI", and argued that the UK needed to be careful not to over-regulate by comparison.⁵⁴⁵
394. However, it was more commonly suggested that it was not plausible to expect the UK to be able to compete, at least in terms of investment, with the US and China.⁵⁴⁶ Dr Pesenti stated this most clearly when he told us that "the UK, because it is smaller, is not going to be the best at everything in AI", but believed there could still be a unique and important role for the UK

541 *Industrial Strategy: Building a Britain fit for the future*, p 36

542 Written evidence from NVIDIA ([AIC0212](#))

543 Written evidence from Balderton Capital (UK) LLP ([AIC0232](#))

544 Written evidence from Microsoft ([AIC0149](#))

545 Written evidence from Ocado Group plc ([AIC0050](#))

546 Written evidence from Will Crosthwait ([AIC0094](#)); The Economic Singularity Supper Club ([AIC0058](#)) and Department of Computer Science University of Liverpool ([AIC0192](#))

on the AI world stage, if it was to be “nimble, clever and move quickly”.⁵⁴⁷ Indeed, we were greatly impressed by the focus and clarity of Canada and Germany’s national strategies when we spoke with Dr Alan Bernstein, President and CEO of CIFAR and Professor Wolfgang Wahlster, CEO and Scientific Director of the DFKI. Dr Bernstein focused on the Pan-Canadian AI Strategy’s bid to attract talented AI developed and researchers back to Canada from the United States, while Professor Wahlster emphasised that Germany was focusing on AI for manufacturing.⁵⁴⁸ We also received evidence from the governments of Japan and the Republic of Korea, informing us of their focus on AI in areas such as manufacturing and robotics, and consumer goods.⁵⁴⁹

395. There are encouraging signs that the UK Government is beginning to think in these terms, and is starting to focus on the concept of ethical AI development and application. Matt Hancock MP was clear that there are “gaps across the world that no one has yet filled”, and it was particularly important to ensure that “we have the structures in place to harness the potential of this technology to improve the lot of humanity”.⁵⁵⁰ The Industrial Strategy reaffirms this stance, stating that “we will lead the world in safe and ethical use of data and artificial intelligence giving confidence and clarity to citizens and business”.⁵⁵¹ In January 2018, the Prime Minister said at the World Economic Forum in Davos that she wanted to establish “the rules and standards that can make the most of artificial intelligence in a responsible way”, and emphasised that the Centre for Data Ethics and Innovation would work with international partners on this project, and that the UK would be joining the World Economic Forum’s new council on artificial intelligence, which aims to help shape global governance in the area.⁵⁵²
396. Indeed, we are also aware of the growing international interest in the governance of AI in recent years—witnesses mentioned the 2016 IEEE AI & Ethics Summit, and the 2017 AI for Good Summit, for example.⁵⁵³ The European Parliament’s interest in this area was also frequently mentioned, and many witnesses were clear that the UK should continue to work with the EU on this area even after Brexit. Thomas Cheney, a researcher in space law at Sunderland University, even called for “a global coordination effort via the United Nations, as was done at the beginning of the Space Age”.⁵⁵⁴ ORBIT brought these points together when they stated:

“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”.⁵⁵⁵

547 [Q 209](#) (Dr Jérôme Pesenti)

548 [Q 173](#) (Dr Alan Bernstein) and [Q 164](#) (Professor Wolfgang Wahlster)

549 Written evidence from Government of Japan ([AIC0224](#)) and Government of the Republic of Korea ([AIC0228](#))

550 [Q 190](#) (Matt Hancock MP)

551 *Industrial Strategy: Building a Britain fit for the future*, p 40

552 Prime Minister Theresa May, Speech at Davos 2018, 25 January 2018: <https://www.gov.uk/government/speeches/pms-speech-at-davos-2018-25-january> [accessed 1 February 2018]

553 Written evidence from IEEE European Public Policy Initiative Working Group on ICT ([AIC0106](#)) and Amnesty International ([AIC0180](#))

554 Written evidence from Mr Thomas Cheney ([AIC0098](#))

555 Written evidence from ORBIT The Observatory for Responsible Research and Innovation in ICT ([AIC0109](#))

They further mentioned the Council of Europe's proposals for close cooperation between themselves, the EU and UNESCO to develop a harmonised legal framework and regulatory mechanisms at the international level. We also received direct evidence of the appetite for greater international co-operation on AI matters. For example, the government of China told us of its hope "to promote closer communication and co-operation" between the UK and China on AI.⁵⁵⁶

397. However, there are also countervailing trends which are less encouraging. When we visited the Leverhulme Centre for the Future of Intelligence, their researchers warned of a potential 'AI arms race' emerging, as various countries seek to develop more sophisticated AI, and potentially disregard concerns around safety and ethics in the process.⁵⁵⁷ Last year, Russian President Vladimir Putin's speech on AI attracted attention worldwide, when he observed that alongside "colossal opportunities" it also brought "threats that are difficult to predict", and that "whoever becomes the leader in this sphere will become the ruler of the world".⁵⁵⁸ However, he also emphasised that, should Russia become a leader, "we will share this know-how with the entire world, the same way we share our nuclear technologies today". A number of witnesses suggested to us that China's relative lack of interest in moderating the use of data by the state and private sector is giving it a competitive advantage relative to more privacy conscious western nations.⁵⁵⁹
398. On the basis of the evidence we have received, we are convinced that vague statements about the UK 'leading' in AI are unrealistic and unhelpful, especially given the vast scale of investment in AI by both the USA and China. By contrast, countries such as Germany and Canada are developing cohesive strategies which take account of their circumstances and seek to play to their strengths as a nation. The UK can either choose to actively define a realistic role for itself with respect to AI, or be relegated to the role of a passive observer.
399. We believe it is very much in the UK's interest to take a lead in steering the development and application of AI in a more co-operative direction, and away from this riskier and ultimately less beneficial vision of a global 'arms race'. The kind of AI-powered future we end up with will ultimately be determined by many countries, whether by collaboration or competition, and whatever the UK decides for itself will ultimately be for naught if the rest of the world moves in a different direction. It is therefore imperative that the Government, and its many internationally-respected institutions, facilitate this global discussion and put forward its own practical ideas for the ethical development and use of AI.
400. We should take advantage of the demand for considered and joined-up ethical principles and frameworks for the development and use of AI in democratic societies. The United States is unlikely to take this role. Not only does the current administration appear relatively uninterested in AI, and has taken a cautious stance on international leadership more generally, the overwhelming dominance of a few powerful technology companies in the

556 Written evidence from the Government of China ([AIC0145](#))

557 See Appendix 5.

558 Written evidence from Information Systems Audit and Control Association (ISACA) London Chapter ([AIC0193](#))

559 Written evidence from Simul Systems Ltd ([AIC0016](#)); Ocado Group plc ([AIC0050](#)) and Will Crosthwait ([AIC0094](#))

development of AI there makes it less likely that a truly democratic debate of equals, encompassing the state, the private sector, universities and the public, is likely to emerge there. Similarly, China shows few signs of wishing to limit the purview of the state or state-supported companies in utilising AI for alarmingly intrusive purposes.

401. The UK therefore has a unique opportunity to forge a distinctive role for itself as a pioneer in ethical AI, which would play to our particular blend of national assets. Alongside a very strong tradition of computer science research in our universities, we also have world-leading humanities departments, who can provide invaluable insight and context regarding the ethical and societal implications of AI. Furthermore, our successful AI start-up sector is enhanced by their close proximity to associated areas of business, most notably our thriving fintech sector,⁵⁶⁰ which can serve as practical testbeds for ethical AI development. We have some of the world's foremost law firms, legal experts and civic institutions, all of which can help enshrine the values and principles we arrive at in robust legal and civic mechanisms where necessary. And finally, we have world-respected institutions such as the BBC, alongside a long history of international diplomacy, engagement and leadership, which will be necessary if we are to help convene, guide and shape the international debates which need to happen in this area.
402. **The transformative potential for artificial intelligence on society at home, and abroad, requires active engagement by one and all. The Government has an opportunity at this point in history to shape the development and deployment of artificial intelligence to the benefit of everyone. The UK's strengths in law, research, financial services and civic institutions, mean it is well placed to help shape the ethical development of artificial intelligence and to do so on the global stage. To be able to demonstrate such influence internationally, the Government must ensure that it is doing everything it can for the UK to maximise the potential of AI for everyone in the country.**
403. **We recommend that the Government convene a global summit in London by the end of 2019, in close conjunction with all interested nations and governments, industry (large and small), academia, and civil society, on as equal a footing as possible. The purpose of the global summit should be to develop a common framework for the ethical development and deployment of artificial intelligence systems. Such a framework should be aligned with existing international governance structures.**

An AI Code

404. While the precise impact of AI across society, politics and the economy remains uncertain, it is generally not disputed that it will have some effect on all three. If poorly handled, public confidence in artificial intelligence could be undermined. The public are entitled to be reassured that AI will be used in their interests, and will not be used to exploit or manipulate them, and many organisations and companies are as eager to confirm these hopes and assuage these concerns.
405. We heard from a number of companies who are developing and publishing their own principles for the ethical development and use of AI, as well as about a number of other ethics-orientated initiatives. In January 2017 the

⁵⁶⁰ Financial technology, or fintech, refers to any technology used to support or enable banking and financial services.

chief executive officer (CEO) of IBM, Ginni Rometty, proposed three core principles for designing and developing AI at IBM, focused on ensuring: that AI is used to augment, rather than replace, human labour; that AI systems are designed to be transparent; and that workers and citizens are properly trained and educated in the use of AI products and services.⁵⁶¹ Sage, who have been developing AI-powered accounting software, announced ‘five core principles’ for developing AI for business in June 2017, which focused on diversity, transparency, accessibility, accountability and augmenting rather than replacing human labour.⁵⁶² SAP, Nvidia and others told us of similar initiatives. DeepMind has developed this one stage further, recently launching their ‘Ethics & Society’ unit, which we were told would help them “explore and understand the real world impacts of AI”, and 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”.⁵⁶³

406. The Market Research Society told us that the “use of ethics boards and ethics reviews committees and processes within a self-regulatory framework will be important tools”.⁵⁶⁴ Eileen Burbidge explained the benefits of this approach to companies, and said “the companies employing AI technology, to the extent they demonstrate they have ethics boards, review their policies and understand their principles, will be the ones to attract the clients, the customers, the partners and the consumers more readily than others that do not or are not as transparent about that”.⁵⁶⁵
407. There are a number of organisations now attempting to devise similar ethical initiatives, often at an international level. Over the course of 2017 the Partnership on AI, an international, pan-industry organisation which aims to bring together researchers, academics, businesses and policymakers, started to take shape. Alongside a number of companies, including Google, Facebook, IBM, Microsoft and Amazon, the Partnership includes a number of university departments and non-governmental organisations (NGOs). It has announced a range of initiatives, including the establishment of a number of working groups to research and formulate best practices, a fellowship program aimed at assisting NGOs, and a series of AI Grand Challenges aimed at using AI to address long-term societal issues.⁵⁶⁶
408. The Institute of Electrical and Electronics Engineers (IEEE) also told us of their efforts to develop a set of internationally accepted ethical principles, with their design manual, *Ethically Aligned Design*, and their IEEE P7000 series of ethically oriented standards.⁵⁶⁷ Closer to home, the British Standards Institution also told us of their similar efforts, which led to the publication of their *Guide to the ethical design and application of robots and robotic systems* (BS 8611:2016).⁵⁶⁸ Most recently, Nesta produced a draft outline of ten

561 Alison DeNisco Rayome, ‘3 guiding principles for ethical AI, from IBM CEO Ginni Rometty’, *TechRepublic* (17 January 2017): <https://www.techrepublic.com/article/3-guiding-principles-for-ethical-ai-from-ibm-ceo-ginni-rometty/> [accessed 5 February 2018]

562 Written evidence from Sage (AIC0159)

563 Written evidence from DeepMind (AIC0234)

564 Written evidence from the Market Research Society (AIC0130)

565 Q 52 (Eileen Burbidge)

566 ‘Partnership on AI strengthens its network of partners and announces first initiatives’, *Partnership on AI* (16 May 2017): <https://www.partnershiponai.org/2017/05/pai-announces-new-partners-and-initiatives/> [accessed 5 February 2018]

567 Written evidence from IEEE European Public Policy Initiative Working Group on ICT (AIC0106)

568 Written evidence from British Standards Institution (AIC0165)

principles for the public sector use of algorithmic decision making.⁵⁶⁹ Finally, the Nuffield Foundation have announced the creation of an independent Convention on Data Ethics and Artificial Intelligence, bringing together various experts to examine ethical issues on a rolling basis, which they intend to convene by the end of 2018.⁵⁷⁰

409. While these efforts are to be encouraged, it was stressed to us that there is still a lack of co-ordination, especially at the national level. Andrew de Rozairo told us that in his view that there was a need for a “multi-stakeholder dialogue” on a national basis in the UK, and pointed to SAP’s engagement with these approaches in France and Germany.⁵⁷¹ Likewise, Kriti Sharma said that while she had taken Sage’s ethical principles to 1,500 AI developers in London, she believed that more needed to be done to ensure that industry shared and collaborated on these principles “because this will not work if it is just Sage, SAP or IBM doing it in silos alone”.⁵⁷² She believed there was a role for Government to help facilitate this collaboration, and help “identify that ultimate checklist and then share it with the industry bodies and have the executives and boards consider that as the things they need to care about”.⁵⁷³
410. There are also questions over how companies will translate these principles into practice, and the degree of accountability which companies and organisations will face if they violate them. For example, when we asked Dr Timothy Lanfear how Nvidia ensured their own workforce was aware of their ethical principles, and how they ensure compliance, he admitted that he struggled to answer the question, because “as a technologist it is not my core thinking”.⁵⁷⁴ It is unlikely Dr Lanfear is alone in this, and mechanisms must be found to ensure the current trend for ethical principles does not simply translate into a meaningless box-ticking exercise. Dr Lynch was altogether more sceptical, arguing “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”.⁵⁷⁵ He believed that only legal frameworks and appropriate regulation would suffice. On the other hand, Eileen Burbidge, took the view that proper ethical governance made good business sense:

“AI companies or the companies employing AI technology, to the extent they demonstrate they have ethics boards, review their policies and understand their principles, will be the ones to attract the clients, the customers, the partners and the consumers more readily than others that do not or are not as transparent about that”.⁵⁷⁶

411. The pace at which the Government has approached these issues has been varied. In some respects, it has been ahead of the curve, and in May 2016 Matt Hancock MP announced the first *Data Science Ethical Framework* for

569 Eddie Copeland, ‘10 principles for public sector use of algorithmic decision making’, *Nesta blog* (21 February 2018): <https://www.nesta.org.uk/code-of-standards-public-sector-use-algorithmic-decision-making> [accessed 1 March 2018]

570 Written evidence from The Alan Turing Institute (AIC0139)

571 Q 79 (Andrew de Rozairo)

572 Q 79 (Kriti Sharma)

573 *Ibid.*

574 Q 43 (Dr Timothy Lanfear)

575 Supplementary written evidence from Dr Mike Lynch (AIC0230)

576 Q 52 (Eileen Burbidge)

public consultation, which outlined six ‘key principles’ intended to guide the work of public sector data scientists:

- Start with clear user need and public benefit;
- Use data and tools which have the minimum intrusion necessary;
- Create robust data science models;
- Be alert to public perceptions;
- Be as open and accountable as possible; and
- Keep data secure.⁵⁷⁷

412. The Framework was developed with advice from the ICO, who confirmed that it could form the basis for Data Protection Impact Assessments (see Box 13), as would be required by the EU’s General Data Protection Regulation (GDPR).⁵⁷⁸ However, since its announcement, its development has lacked a sense of urgency, with very little reference to it in the intervening years. Dr Jerry Fishenden, an expert on digital government, while welcoming the original draft, noted in July 2017 that “since the launch it’s unclear what the status of the framework is. There are no indications of any consultation taking place, or resulting improvements, on the website ... the execution since its launch lacks credibility”.⁵⁷⁹

Box 13: Data Protection Impact Assessments

Data protection impact assessments (DPIAs) is a new requirement of the GDPR (and likely the Data Protection Bill). DPIAs are intended as a tool to help organisations identify the most effective way to comply with their data protection obligations and meet individuals’ expectations of privacy. According to the ICO, one must carry out a DPIA when using new technologies and the processing is likely to result in a high risk to the rights and freedoms of individuals. This means organisations and companies which choose to deploy AI systems in the near future will likely have to produce them. A DPIA should include:

- A description of the processing operations and the purposes, including, where applicable, the legitimate interests pursued by the controller.
- An assessment of the necessity and proportionality of the processing in relation to the purpose.
- An assessment of the risks to individuals.
- The measures in place to address risk, including security and to demonstrate that you comply.

Source: ICO, ‘Data protection impact assessments’: <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments> [accessed 1 February 2018]

577 Cabinet Office, *Data Science Ethical Framework* (19 May 2016): https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication_1.pdf [accessed 31 January 2018]

578 *Ibid.*

579 Dr Jerry Fishenden, ‘Improving data science ethics’, *New tech observations from the UK* (5 July 2017): <https://ntouk.wordpress.com/2017/07/05/improving-data-science-ethics/> [accessed 31 January 2018]

413. More recently a number of announcements have been made in this area. In November 2017 the Government Digital Service announced they were updating the Framework, based on feedback from the British Academy, the Royal Society and Essex County Council.⁵⁸⁰ The Government also announced the creation of the Centre for Data Ethics and Innovation, described as aiming “to enable and ensure safe, ethical and ground-breaking innovation in AI and data-driven technologies”.⁵⁸¹ This “world-first advisory body” will work with “Government, regulators and industry to lay the foundations for AI adoption”.⁵⁸² Finally, in January 2018, there was also the creation of a new *Digital Charter*, which the Government has described as a “rolling programme of work to agree norms and rules for the online world and put them into practice”, which would aim to ensure that people have “the same rights and expect the same behaviour online as [they] do offline”.⁵⁸³ ‘Data and artificial intelligence ethics and innovation’ will constitute one of the seven elements of this work programme.⁵⁸⁴
414. From all we have seen, we believe this area is not lacking good will, but there is a lack of awareness and co-ordination, which is where Government involvement could help. It is also clear to us that this is not only an ethical matter, but also good business sense. The evidence we have received suggests that some individuals, organisations and public services are reluctant to share data with companies because they do not know what is acceptable—a particular concern after the Royal Free London NHS Foundation Trust’s deal with DeepMind (see Box 8).⁵⁸⁵
415. A core set of widely recognised ethical principles, which companies and organisations deploying AI sign up to in the form of a code, could be useful in this context. The Digital Charter may yet turn into this, although given the slow development of the Government’s Data Science Ethical Framework, there are reasons to be sceptical. Nevertheless, whether it is positioned within the broader framework of the Digital Charter or independent of it, there is a need for clear and understandable guidelines governing the applications to which AI may be put, between businesses, public organisations and individual consumers.
416. These guidelines should be developed with substantive input from the Centre for Data Ethics and Innovation, the AI Council and the Alan Turing Institute. It should include both organisation-level considerations, as well as questions and checklists for those designing, developing and utilising AI at an operational level, alongside concrete examples of how this should work in practice.

580 Sarah Gates, ‘Updating the Data Science Ethical Framework’, *Government Digital Service blog* (27 November 2017): <https://gds.blog.gov.uk/2017/11/27/updating-the-data-science-ethical-framework/> [accessed 31 January 2018]

581 *Autumn Budget 2017*, p 4

582 *Ibid.*

583 Department for Digital, Culture, Media and Sport, *Digital Charter* (25 January 2018): <https://www.gov.uk/government/publications/digital-charter> [accessed 31 January 2018]

584 *Ibid.*

585 Written evidence from medConfidential (AIC0063); Future Advocacy (AIC0121); Doteveryone (AIC0148) and Royal Statistical Society (AIC0218)

417. As a starting point in this process, we suggest five overarching principles for an AI Code:
- (1) Artificial intelligence should be developed for the common good and benefit of humanity.
 - (2) Artificial intelligence should operate on principles of intelligibility and fairness.
 - (3) Artificial intelligence should not be used to diminish the data rights or privacy of individuals, families or communities.
 - (4) All citizens have the right to be educated to enable them to flourish mentally, emotionally and economically alongside artificial intelligence.
 - (5) The autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence.
418. Furthermore, while we do not see this having a statutory basis, at least initially, consumers in particular should be able to trust that someone external to the companies and organisations which adopt these principles has some measure of oversight regarding their adherence to them. An appropriate organisation, such as the Centre for Data Ethics and Innovation, could be assigned to oversee the adherence of signed-up organisations and companies to this code, and offer advice on how to improve where necessary. In more extreme cases, they may even consider withdrawing this seal of approval. To organisations and businesses, it would provide a clear, consistent and interoperable framework for their activities, while for citizens and consumers, it would provide a recognised and trustworthy brand, reassuring them across the multiple domains of their life that they were getting a fair deal from AI.
419. **Many organisations are preparing their own ethical codes of conduct for the use of AI. This work is to be commended, but it is clear that there is a lack of wider awareness and co-ordination, where the Government could help. Consistent and widely-recognised ethical guidance, which companies and organisations deploying AI could sign up to, would be a welcome development.**
420. **We recommend that a cross-sector ethical code of conduct, or ‘AI code’, suitable for implementation across public and private sector organisations which are developing or adopting AI, be drawn up and promoted by the Centre for Data Ethics and Innovation, with input from the AI Council and the Alan Turing Institute, with a degree of urgency. In some cases, sector-specific variations will need to be created, using similar language and branding. Such a code should include the need to have considered the establishment of ethical advisory boards in companies or organisations which are developing, or using, AI in their work. In time, the AI code could provide the basis for statutory regulation, if and when this is determined to be necessary.**

SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

Engaging with artificial intelligence

General understanding, engagement and public narratives

1. The media provides extensive and important coverage of artificial intelligence, which occasionally can be sensationalist. It is not for the Government or other public organisations to intervene directly in how AI is reported on, nor to attempt to promote an entirely positive view among the general public of its possible implications or impact. Instead, the Government must understand the need to build public trust and confidence in how to use artificial intelligence, as well as explain the risks. (Paragraph 50)

Everyday engagement with AI

2. Artificial intelligence is a growing part of many people's lives and businesses. It is important that members of the public are aware of how and when artificial intelligence is being used to make decisions about them, and what implications this will have for them personally. This clarity, and greater digital understanding, will help the public experience the advantages of AI, as well as to opt out of using such products should they have concerns. (Paragraph 58)
3. Industry should take the lead in establishing voluntary mechanisms for informing the public when artificial intelligence is being used for significant or sensitive decisions in relation to consumers. This industry-led approach should learn lessons from the largely ineffective AdChoices scheme. The soon-to-be established AI Council, the proposed industry body for AI, should consider how best to develop and introduce these mechanisms. (Paragraph 59)

Designing artificial intelligence

Access to, and control of, data

4. The Government plans to adopt the Hall-Pesenti Review recommendation that 'data trusts' be established to facilitate the ethical sharing of data between organisations. However, under the current proposals, individuals who have their personal data contained within these trusts would have no means by which they could make their views heard, or shape the decisions of these trusts. We therefore recommend that as data trusts are developed under the guidance of the Centre for Data Ethics and Innovation, provision should be made for the representation of people whose data is stored, whether this be via processes of regular consultation, personal data representatives, or other means. (Paragraph 82)
5. Access to data is essential to the present surge in AI technology, and there are many arguments to be made for opening up data sources, especially in the public sector, in a fair and ethical way. Although a 'one-size-fits-all' approach to the handling of public sector data is not appropriate, many SMEs in particular are struggling to gain access to large, high-quality datasets, making it extremely difficult for them to compete with the large, mostly US-owned technology companies, who can purchase data more easily and are also large enough to generate their own. In many cases, public datasets, such as those held by the NHS, are more likely to contain data on more diverse populations than their private sector equivalents, and more control can be exercised before they are released. (Paragraph 83)

6. We recommend that wherever possible and appropriate, and with regard to its potential commercial value, publicly-held data be made available to AI researchers and developers. In many cases, this will require Government departments and public organisations making a concerted effort to digitise their records in unified and compatible formats. When releasing this data, subject to appropriate anonymisation measures where necessary, data trusts will play an important role. (Paragraph 84)
7. We support the approach taken by Transport for London, who have released their data through a single point of access, where the data is available subject to appropriate terms and conditions and with controls on privacy. The Centre for Data Ethics and Innovation should produce guidance on similar approaches. The Government Office for AI and GovTech Catalyst should work together to ensure that the data for which there is demand is made available in a responsible manner. (Paragraph 85)
8. We acknowledge that open data cannot be the last word in making data more widely available and usable, and can often be too blunt an instrument for facilitating the sharing of more sensitive or valuable data. Legal and technical mechanisms for strengthening personal control over data, and preserving privacy, will become increasingly important as AI becomes more widespread through society. Mechanisms for enabling individual data portability, such as the Open Banking initiative, and data sharing concepts such as data trusts, will spur the creation of other innovative and context-appropriate tools, eventually forming a broad spectrum of options between total data openness and total data privacy. (Paragraph 86)
9. We recommend that the Centre for Data Ethics and Innovation investigate the Open Banking model, and other data portability initiatives, as a matter of urgency, with a view to establishing similar standardised frameworks for the secure sharing of personal data beyond finance. They should also work to create, and incentivise the creation of, alternative tools and frameworks for data sharing, control and privacy for use in a wide variety of situations and contexts. (Paragraph 87)
10. Increasingly, public sector data has value. It is important that public organisations are aware of the commercial potential of such data. We recommend that the Information Commissioner's Office work closely with the Centre for Data Ethics and Innovation in the establishment of data trusts, and help to prepare advice and guidance for data controllers in the public sector to enable them to estimate the value of the data they hold, in order to make best use of it and negotiate fair and evidence-based agreements with private-sector partners. The values contained in this guidance could be based on precedents where public data has been made available and subsequently generated commercial value for public good. The Information Commissioner's Office should have powers to review the terms of significant data supply agreements being contemplated by public bodies. (Paragraph 88)

Intelligible AI

11. Based on the evidence we have received, we believe that achieving full technical transparency is difficult, and possibly even impossible, for certain kinds of AI systems in use today, and would in any case not be appropriate or helpful in many cases. However, there will be particular safety-critical scenarios where technical transparency is imperative, and regulators in those domains must have the power to mandate the use of more transparent forms of AI, even at the potential expense of power and accuracy. (Paragraph 99)
12. We believe that the development of intelligible AI systems is a fundamental necessity if AI is to become an integral and trusted tool in our society. Whether this takes the form of technical transparency, explainability, or indeed both, will depend on the context and the stakes involved, but in most cases we believe explainability will be a more useful approach for the citizen and the consumer. This approach is also reflected in new EU and UK legislation. We believe it is not acceptable to deploy any artificial intelligence system which could have a substantial impact on an individual's life, unless it can generate a full and satisfactory explanation for the decisions it will take. In cases such as deep neural networks, where it is not yet possible to generate thorough explanations for the decisions that are made, this may mean delaying their deployment for particular uses until alternative solutions are found. (Paragraph 105)
13. The Centre for Data Ethics and Innovation, in consultation with the Alan Turing Institute, the Institute of Electrical and Electronics Engineers, the British Standards Institute and other expert bodies, should produce guidance on the requirement for AI systems to be intelligible. The AI development sector should seek to adopt such guidance and to agree upon standards relevant to the sectors within which they work, under the auspices of the AI Council. (Paragraph 106)

Addressing prejudice

14. We are concerned that many of the datasets currently being used to train AI systems are poorly representative of the wider population, and AI systems which learn from this data may well make unfair decisions which reflect the wider prejudices of societies past and present. While many researchers, organisations and companies developing AI are aware of these issues, and are starting to take measures to address them, more needs to be done to ensure that data is truly representative of diverse populations, and does not further perpetuate societal inequalities. (Paragraph 119)
15. Researchers and developers need a more developed understanding of these issues. In particular, they need to ensure that data is pre-processed to ensure it is balanced and representative wherever possible, that their teams are diverse and representative of wider society, and that the production of data engages all parts of society. Alongside questions of data bias, researchers and developers need to consider biases embedded in the algorithms themselves—human developers set the parameters for machine learning algorithms, and the choices they make will intrinsically reflect the developers' beliefs, assumptions and prejudices. The main ways to address these kinds of biases are to ensure that developers are drawn from diverse gender, ethnic and socio-economic backgrounds, and are aware of, and adhere to, ethical codes of conduct. (Paragraph 120)

16. We recommend that a specific challenge be established within the Industrial Strategy Challenge Fund to stimulate the creation of authoritative tools and systems for auditing and testing training datasets to ensure they are representative of diverse populations, and to ensure that when used to train AI systems they are unlikely to lead to prejudicial decisions. This challenge should be established immediately, and encourage applications by spring 2019. Industry must then be encouraged to deploy the tools which are developed and could, in time, be regulated to do so. (Paragraph 121)

Data monopolies

17. While we welcome the investments made by large overseas technology companies in the UK economy, and the benefits they bring, the increasing consolidation of power and influence by a select few risks damaging the continuation, and development, of the UK's thriving home-grown AI start-up sector. The monopolisation of data demonstrates the need for strong ethical, data protection and competition frameworks in the UK, and for continued vigilance from the regulators. We urge the Government, and the Competition and Markets Authority, to review proactively the use and potential monopolisation of data by the big technology companies operating in the UK. (Paragraph 129)

Developing artificial intelligence

Investment in AI development

18. The UK AI development sector has flourished largely without attempts by the Government to determine its shape or direction. This has resulted in a flexible and innovative grassroots start-up culture, which is well positioned to take advantage of the unpredictable opportunities that could be afforded by AI. The investment environment for AI businesses must be able to cope with this uncertainty, and be willing to take the risks required to seize the chances AI offers. (Paragraph 135)
19. We welcome the changes announced in the Autumn Budget 2017 to the Enterprise Investment and Venture Capital Trust schemes which encourage innovative growth, and we believe they should help to boost investment in UK-based AI companies. The challenge for start-ups in the UK is the lack of investment available with which to scale up their business. (Paragraph 150)
20. To ensure that AI start-ups in the United Kingdom have the opportunity to scale up, without having to look for off-shore investment, we recommend that a proportion of the £2.5 billion investment fund at the British Business Bank, announced in the Autumn Budget 2017, be reserved as an AI growth fund for SMEs with a substantive AI component, and be specifically targeted at enabling such companies to scale up. Further, the Government should consult on the need to improve access to funding within the UK for SMEs with a substantive AI component looking to scale their business. (Paragraph 151)
21. To guarantee that companies developing AI can continue to thrive in the UK, we recommend that the Government review the existing incentives for businesses operating in the UK who are working on artificial intelligence products, and ensure that they are adequate, properly promoted to companies, and designed to assist SMEs wherever possible. (Paragraph 152)

Turning academic research into commercial potential

22. The UK has an excellent track record of academic research in the field of artificial intelligence, but there is a long-standing issue with converting such research into commercially viable products. (Paragraph 159)
23. To address this we welcome, and strongly endorse, the recommendation of the Hall-Pesenti Review, which stated “universities should use clear, accessible and where possible common policies and practices for licensing IP and forming spin-out companies”. We recommend that the Alan Turing Institute, as the National Centre for AI Research, should develop this concept into concrete policy advice for universities in the UK, looking to examples from other fields and from other nations, to help start to address this long-standing problem. (Paragraph 160)

Improving access to skilled AI developers

24. We welcome the expanded public funding for PhD places in AI and machine learning, as well as the announcement that an industry-funded master’s degree programme is to be developed. We do believe that more needs to be done to ensure that the UK has the pipeline of skills it requires to maintain its position as one of the best countries in the world for AI research. (Paragraph 168)
25. We recommend that the funding for PhD places in AI and machine learning be further expanded, with the financial burden shared equally between the public and private sector through a PhD matching scheme. We believe that the Doctoral Training Partnership scheme and other schemes where costs are shared between the private sector, universities and research councils should be examined, and the number of industry-sponsored PhDs increased. (Paragraph 169)
26. We further recommend that short (3–6 months) post-graduate conversion courses be developed by the Alan Turing Institute, in conjunction with the AI Council, to reflect the needs of the AI development sector. Such courses should be suitable for individuals in other academic disciplines looking to transfer into AI development and design or to have a grounding in the application of AI in their discipline. These should be designed so as to enable anyone to retrain at any stage of their working lives. (Paragraph 170)
27. We recommend that the Government ensures that publically-funded PhDs in AI and machine learning are made available to a diverse population, more representative of wider society. To achieve this, we call for the Alan Turing Institute and Government Office for AI to devise mechanisms to attract more female and ethnic minority students from academic disciplines which require similar skillsets, but have more representative student populations, to participate in the Government-backed PhD programme. (Paragraph 174)
28. We acknowledge the considerable scepticism of at least some technology companies who believe that the apprenticeship levy is of little use to them, despite the success that others in the sector have had with apprenticeships. The Government should produce clear guidance on how the apprenticeship levy can be best deployed for use in the technology sector, in particular in SMEs and start-ups. (Paragraph 175)

29. The Government's announcement that it will increase the annual number of Tier 1 (exceptional talent) visas from 1,000 to 2,000 per year is welcome. While top-tier PhD researchers and designers are required, a thriving AI development sector is also dependent on access to those able to implement artificial intelligence research, whose occupations may fall short of the exceptional talent requirements. (Paragraph 181)
30. We are concerned that the number of workers provided for under the Tier 1 (exceptional talent) visa scheme will be insufficient and the requirements too high level for the needs of UK companies and start-ups. We recommend that the number of visas available for AI researchers and developers be increased by, for example, adding machine learning and associated skills to the Tier 2 Shortage Occupations List. (Paragraph 182)

Maintaining innovation

31. We believe that the Government must commit to underwriting, and where necessary replacing, funding for European research and innovation programmes, after we have left the European Union. (Paragraph 188)
32. The state has an important role in supporting AI research through the research councils and other mechanisms, and should be mindful to ensure that the UK's advantages in AI R&D are maintained. There is a risk that the current focus on deep learning is distracting attention away from other aspects of AI research, which could contribute to the next big advances in the field. The Government and universities have an important role to play in supporting diverse sub-fields of AI research, beyond the now well-funded area of deep learning, in order to ensure that the UK remains at the cutting edge of AI developments. (Paragraph 191)

Working with artificial intelligence

Productivity

33. We support the Government's belief that artificial intelligence offers an opportunity to improve productivity. However, to meet this potential for the UK as a whole, the AI Council must take a role in enabling AI to benefit all companies (big and small) and ensuring they are able to take advantage of existing technology, in order for them to take advantage of future technology. It will be important that the Council identifies accelerators and obstacles to the use of AI to improve productivity, and advises the Government on the appropriate course of action to take. (Paragraph 199)
34. We welcome the Government's intentions to upgrade the nation's digital infrastructure, as far as they go. However, we are concerned that it does not have enough impetus behind it to ensure that the digital foundations of the country are in place in time to take advantage of the potential artificial intelligence offers. We urge the Government to consider further substantial public investment to ensure that everywhere in the UK is included within the rollout of 5G and ultrafast broadband, as this should be seen as a necessity. (Paragraph 203)

Government adoption, and procurement, of artificial intelligence

35. The Government's leadership in the development and deployment of artificial intelligence must be accompanied by action. We welcome the announcement of the GovTech Catalyst and hope that it can open the doors of Whitehall to the burgeoning AI development sector in the UK. We also endorse the recommendation of the Hall-Pesenti Review aimed at encouraging greater use of AI in the public sector. (Paragraph 215)
36. To ensure greater uptake of AI in the public sector, and to lever the Government's position as a customer in the UK, we recommend that public procurement regulations are reviewed and amended to ensure that UK-based companies offering AI solutions are invited to tender and given the greatest opportunity to participate. The Crown Commercial Service, in conjunction with the Government Digital Office, should review the Government Service Design Manual and the Technology Code of Practice to ensure that the procurement of AI-powered systems designed by UK companies is encouraged and incentivised, and done in an ethical manner. (Paragraph 216)
37. We also encourage the Government to be bold in its approach to the procurement of artificial intelligence systems, and to encourage the development of possible solutions to public policy challenges through limited speculative investment and support to businesses which helps them convert ideas to prototypes, in order to determine whether their solutions are viable. The value of AI systems which are deployed to the taxpayer will compensate for any money lost in supporting the development of other tools. (Paragraph 217)
38. Finally, with respect to public procurement, we recommend the establishment of an online bulletin board for the advertisement of challenges which the Government Office for AI and the GovTech Catalyst have identified from across Government and the wider public sector where there could be the potential for innovative tech- and AI-based solutions. (Paragraph 218)

Impact on the labour market

39. The labour market is changing, and further significant disruption to that market is expected as AI is adopted throughout the economy. As we move into this unknown territory, forecasts of AI's growing impact—jobs lost, jobs enhanced and new jobs created—are inevitably speculative. There is an urgent need to analyse or assess, on an ongoing basis, the evolution of AI in the UK, and develop policy responses. (Paragraph 231)

National Retraining Scheme

40. The UK must be ready for the disruption that AI will have on the way in which we work. We support the Government's interest in developing adult retraining schemes, as we believe that AI will disrupt a wide range of jobs over the coming decades, and both blue- and white-collar jobs which exist today will be put at risk. It will therefore be important to encourage and support workers as they move into the new jobs and professions we believe will be created as a result of new technologies, including AI. The National Retraining Scheme could play an important role here, and must ensure that the recipients of retraining schemes are representative of the wider population. Industry should assist in the financing of the National Retraining Scheme by matching Government funding. This partnership would help improve

the number of people who can access the scheme and better identify the skills required. Such an approach must reflect the lessons learned from the execution of the Apprenticeship Levy. (Paragraph 236)

Living with artificial intelligence

Education and artificial intelligence

41. It is clear to us that there is a need to improve digital understanding and data literacy across society, as these are the foundations upon which knowledge about AI is built. This effort must be undertaken collaboratively by public sector organisations, civil society organisations (such as the Royal Society) and the private sector. (Paragraph 249)
42. The evidence suggests that recent reforms to the computing curriculum are a significant improvement on the ICT curriculum, although it is still too early to say what the final results of this will be. The Government must be careful not to expand computing education at the expense of arts and humanities subjects, which hone the creative, contextual and analytical skills which will likely become more, not less, important in a world shaped by AI. (Paragraph 250)
43. We are, however, concerned to learn of the absence of wider social and ethical implications from the computing curriculum, as originally proposed. We recommend that throughout the curriculum the wider social and ethical aspects of computer science and artificial intelligence need to be restored to the form originally proposed. (Paragraph 251)
44. While we welcome the measures announced in the Autumn Budget 2017 to increase the number of computer science teachers in secondary schools, a greater sense of urgency and commitment is needed from the Government if the UK is to meet the challenges presented by AI. (Paragraph 257)
45. The Government must ensure that the National Centre for Computing is rapidly created and adequately resourced, and that there is support for the retraining of teachers with associated skills and subjects such as mathematics. In particular, Ofsted should ensure that schools are making additional time available to teachers to enable them to train in new technology-focused aspects of the curriculum. We also urge the Government to make maximum use across the country of existing lifelong learning facilities for the training and regular retraining of teachers and other AI experts. (Paragraph 258)
46. Supplementary to the Hall-Pesenti Review, the Government should explore ways in which the education sector, at every level, can play a role in translating the benefits of AI into a more productive and equitable economy. (Paragraph 259)

Impact on social and political cohesion

47. There are many social and political impacts which AI may have, quite aside from people's lives as workers and consumers. AI makes the processing and manipulating of all forms of digital data substantially easier, and given that digital data permeates so many aspects of modern life, this presents both opportunities and unprecedented challenges. As discussed earlier in our report, there is a rapidly growing need for public understanding of, and engagement with, AI to develop alongside the technology itself. The manipulation of data in particular will be a key area for public understanding and discussion in the coming months and years. (Paragraph 265)

48. We recommend that the Government and Ofcom commission research into the possible impact of AI on conventional and social media outlets, and investigate measures which might counteract the use of AI to mislead or distort public opinion as a matter of urgency. (Paragraph 266)

Inequality

49. The risk of greater societal and regional inequalities emerging as a consequence of the adoption of AI and advances in automation is very real, and while the Government's proposed policies on regional development are to be welcomed, we believe more needs to be done in this area. We are not yet convinced that basic income schemes will prove to be the answer, but we watch Scotland's experiments with interest. (Paragraph 275)
50. Everyone must have access to the opportunities provided by AI. The Government must outline its plans to tackle any potential societal or regional inequality caused by AI, and this must be explicitly addressed as part of the implementation of the Industrial Strategy. The Social Mobility Commission's annual State of the Nation report should include the potential impact of AI and automation on inequality. (Paragraph 276)

Healthcare and artificial intelligence

51. Maintaining public trust over the safe and secure use of their data is paramount to the successful widespread deployment of AI and there is no better exemplar of this than personal health data. There must be no repeat of the controversy which arose between the Royal Free London NHS Foundation Trust and DeepMind. If there is, the benefits of deploying AI in the NHS will not be adopted or its benefits realised, and innovation could be stifled. (Paragraph 300)
52. The data held by the NHS could be considered a unique source of value for the nation. It should not be shared lightly, but when it is, it should be done in a manner which allows for that value to be recouped. We are concerned that the current piecemeal approach taken by NHS Trusts, whereby local deals are struck between AI developers and hospitals, risks the inadvertent under-appreciation of the data. It also risks NHS Trusts exposing themselves to inadequate data sharing arrangements. (Paragraph 301)
53. We recommend that a framework for the sharing of NHS data should be prepared and published by the end of 2018 by NHS England (specifically NHS Digital) and the National Data Guardian for Health and Care. This should be prepared with the support of the ICO and the clinicians and NHS Trusts which already have experience of such arrangements (such as the Royal Free London and Moorfields Eye Hospital NHS Foundation Trusts), as well as the Caldicott Guardians. This framework should set out clearly the considerations needed when sharing patient data in an appropriately anonymised form, the precautions needed when doing so, and an awareness of the value of that data and how it is used. It must also take account of the need to ensure SME access to NHS data, and ensure that patients are made aware of the use of their data and given the option to opt out. (Paragraph 302)

54. Many organisations in the United Kingdom are not taking advantage of existing technology, let alone ready to take advantage of new technology such as artificial intelligence. The NHS is, perhaps, the most pressing example of this. The development, and eventual deployment, of AI systems in healthcare in the UK should be seen as a collaborative effort with both the NHS and the AI developer being able to benefit. To release the value of the data held, we urge the NHS to digitise its current practices and records, in consistent formats, by 2022 to ensure that the data it holds does not remain inaccessible and the possible benefits to society unrealised. (Paragraph 303)

Mitigating the risks of artificial intelligence

Legal liability

55. In our opinion, it is possible to foresee a scenario where AI systems may malfunction, underperform or otherwise make erroneous decisions which cause harm. In particular, this might happen when an algorithm learns and evolves of its own accord. It was not clear to us, nor to our witnesses, whether new mechanisms for legal liability and redress in such situations are required, or whether existing mechanisms are sufficient. (Paragraph 317)
56. Clarity is required. We recommend that the Law Commission consider the adequacy of existing legislation to address the legal liability issues of AI and, where appropriate, recommend to Government appropriate remedies to ensure that the law is clear in this area. At the very least, this work should establish clear principles for accountability and intelligibility. This work should be completed as soon as possible. (Paragraph 318)

Criminal misuse of artificial intelligence and data

57. The potential for well-meaning AI research to be used by others to cause harm is significant. AI researchers and developers must be alive to the potential ethical implications of their work. The Centre for Data Ethics and Innovation and the Alan Turing Institute are well placed to advise researchers on the potential implications of their work, and the steps they can take to ensure that such work is not misused. However, we believe additional measures are required. (Paragraph 328)
58. We recommend that universities and research councils providing grants and funding to AI researchers must insist that applications for such money demonstrate an awareness of the implications of the research and how it might be misused, and include details of the steps that will be taken to prevent such misuse, before any funding is provided. (Paragraph 329)
59. We recommend that the Cabinet Office's final Cyber Security Science & Technology Strategy take into account the risks as well as the opportunities of using AI in cybersecurity applications, and applications more broadly. In particular, further research should be conducted into methods for protecting public and private datasets against any attempts at data sabotage, and the results of this research should be turned into relevant guidance. (Paragraph 333)

Autonomous weapons

60. Without agreed definitions we could easily find ourselves stumbling through a semantic haze into dangerous territory. The Government's definition of an autonomous system used by the military as one where it "is capable of understanding higher-level intent and direction" is clearly out of step with the definitions used by most other governments. This position limits both the extent to which the UK can meaningfully participate in international debates on autonomous weapons and its ability to take an active role as a moral and ethical leader on the global stage in this area. Fundamentally, it also hamstrings attempts to arrive at an internationally agreed definition. (Paragraph 345)
61. We recommend that the UK's definition of autonomous weapons should be realigned to be the same, or similar, as that used by the rest of the world. To produce this definition the Government should convene a panel of military and AI experts to agree a revised form of words. This should be done within eight months of the publication of this report. (Paragraph 346)

*Shaping artificial intelligence**Leading at home*

62. Artificial intelligence's potential is an opportunity the Government is embracing. The Government's recent enthusiasm and responsiveness to artificial intelligence in the UK is to be welcomed. We have proposed a number of recommendations for strengthening recent policy announcements, based on the extensive evidence we have received as a Committee. We encourage the Government to continue to be proactive in developing policies to harness the potential of AI and mitigate the risks. We do, however, urge the Government to ensure that its approach is focused and that it provides strategic leadership—there must be a clear roadmap for success. Policies must be produced in concert with one another, and with existing policy. Industry and the public must be better informed about the announcements, and sufficient detail provided from the outset. (Paragraph 366)
63. The pace at which this technology will grow is unpredictable, and the policy initiatives have been many. To avoid policy being too reactive, and to prevent the new institutions from overlapping and conflicting with one another, we recommend that the Government Office for AI develop a national policy framework for AI, to be in lockstep with the Industrial Strategy, and to be overseen by the AI Council. Such a framework should include policies related to the recommendations of this report, and be accompanied, where appropriate, by a long-term commitment to such policies in order to realise the benefits. It must also be clear within Government who is responsible around the Cabinet table for the direction and ownership of this framework and the AI-related policies which fall within it. (Paragraph 367)
64. The roles and remits of the new institutions must be clear, if they are to be a success. The public and the technology sector in the UK must know who to turn to for authoritative advice when it comes to the development and use of artificial intelligence. To ensure public confidence, it must also be clear who to turn to if there are any complaints about how AI has been used, above and beyond the matters relating to data use (which falls within the Information Commissioner's remit). (Paragraph 368)

65. We recommend that the Government Office for AI should act as the co-ordinator of the work between the Centre for Data Ethics and Innovation, the GovTech Catalyst team and the national research centre for Artificial Intelligence Research (the Alan Turing Institute), as well as the AI Council it is being established to support. It must also take heed of the work of the more established bodies which have done work in this area, such as the Information Commissioner's Office and the Competition and Markets Authority. The work programmes of all the new AI-specific institutions should be subject to agreement with one another, on a quarterly basis, and should take into account the work taking place across Government in this area, as well as the recommendations from Parliament, regulators, and the work of the devolved assemblies and governments. The UK has a thriving AI ecosystem, and the Government Office for AI should seek to inform its work programme through wide public consultation as it develops Government policy with regard to artificial intelligence. The programme should be publicly available for scrutiny. (Paragraph 369)
66. We welcome the new focus for the Alan Turing Institute as the national research centre for artificial intelligence. We want it to be able to fulfil this role, and believe it has the potential to do so. As such, the new focus must not simply be a matter of rebranding. The successful institutes in Canada and Germany, such as the Vector Institute and the German Research Center for Artificial Intelligence, offer valuable lessons as to how a national research centre should be operated. (Paragraph 370)
67. The Government must ensure that the Alan Turing Institute's funding and structure is sufficient for it to meet its new expanded remit as the UK's national research centre for AI. In particular, the Institute's current secondment-based staffing model should be assessed to ensure its suitability, and steps taken to staff the Institute appropriately to meet the demands now placed upon it. (Paragraph 371)

Regulation and regulators

68. Blanket AI-specific regulation, at this stage, would be inappropriate. We believe that existing sector-specific regulators are best placed to consider the impact on their sectors of any subsequent regulation which may be needed. We welcome that the Data Protection Bill and GDPR appear to address many of the concerns of our witnesses regarding the handling of personal data, which is key to the development of AI. The Government Office for AI, with the Centre for Data Ethics and Innovation, needs to identify the gaps, if any, where existing regulation may not be adequate. The Government Office for AI must also ensure that the existing regulators' expertise is utilised in informing any potential regulation that may be required in the future and we welcome the introduction of the Regulator's Pioneer Fund. (Paragraph 386)
69. The additional burden this could place on existing regulators may be substantial. We recommend that the National Audit Office's advice is sought to ensure that existing regulators, in particular the Information Commissioner's Office, are adequately and sustainably resourced. (Paragraph 387)

Assessing policy outcomes

70. It is essential that policies towards artificial intelligence are suitable for the rapidly changing environment which they are meant to support. For the UK to be able to realise the benefits of AI, the Government's policies, underpinned by a co-ordinated approach, must be closely monitored and react to feedback from academia and industry where appropriate. Policies should be benchmarked and tracked against appropriate international comparators. The Government Office for AI has a clear role to play here, and we recommend that progress against the recommendations of this report, the Government's AI-specific policies within the Industrial Strategy and other related policies, be reported on an annual basis to Parliament. (Paragraph 391)

A vision for Britain in an AI world

71. The transformative potential for artificial intelligence on society at home, and abroad, requires active engagement by one and all. The Government has an opportunity at this point in history to shape the development and deployment of artificial intelligence to the benefit of everyone. The UK's strengths in law, research, financial services and civic institutions, mean it is well placed to help shape the ethical development of artificial intelligence and to do so on the global stage. To be able to demonstrate such influence internationally, the Government must ensure that it is doing everything it can for the UK to maximise the potential of AI for everyone in the country. (Paragraph 402)
72. We recommend that the Government convene a global summit in London by the end of 2019, in close conjunction with all interested nations and governments, industry (large and small), academia, and civil society, on as equal a footing as possible. The purpose of the global summit should be to develop a common framework for the ethical development and deployment of artificial intelligence systems. Such a framework should be aligned with existing international governance structures. (Paragraph 403)

An AI Code

73. Many organisations are preparing their own ethical codes of conduct for the use of AI. This work is to be commended, but it is clear that there is a lack of wider awareness and co-ordination, where the Government could help. Consistent and widely-recognised ethical guidance, which companies and organisations deploying AI could sign up to, would be a welcome development. (Paragraph 419)
74. We recommend that a cross-sector ethical code of conduct, or 'AI code', suitable for implementation across public and private sector organisations which are developing or adopting AI, be drawn up and promoted by the Centre for Data Ethics and Innovation, with input from the AI Council and the Alan Turing Institute, with a degree of urgency. In some cases, sector-specific variations will need to be created, using similar language and branding. Such a code should include the need to have considered the establishment of ethical advisory boards in companies or organisations which are developing, or using, AI in their work. In time, the AI code could provide the basis for statutory regulation, if and when this is determined to be necessary. (Paragraph 420)

APPENDIX 1: LIST OF MEMBERS AND DECLARATIONS OF INTEREST

Members

Baroness Bakewell
 Lord Clement-Jones (Chairman)
 Lord Giddens
 Baroness Greender
 Lord Hollick
 Lord Holmes of Richmond
 Lord Levene of Portsoken
 The Lord Bishop of Oxford
 Lord Puttnam
 Viscount Ridley
 Baroness Rock
 Lord St John of Bletso
 Lord Swinfen

Declarations of interest

Baroness Bakewell
No relevant interests declared.

Lord Clement-Jones
Chair of the Board, Ombudsman Services Limited
Partner, DLA Piper UK LLP
Chair of the Council of Queen Mary University of London
Co-Chair, All-Party Parliamentary Group on Artificial Intelligence

Lord Giddens
No relevant interests declared.

Baroness Greender
No relevant interests declared.

Lord Hollick
Director and shareholder, Honeywell International Inc
Senior Adviser and shareholder, We Predict
Senior Adviser and shareholder, GP Bullhound Ltd
Advisor and shareholder, Brands Eye
Member, Advisory Board, Royal Society

Lord Holmes of Richmond
Chair, GDI Hub (Global Disability Innovation Hub)
Speaking engagement, 26 July 2017, Apple, London
Microsoft HoloLens demonstration, Guide Dogs for the Blind, 25 July 2017, London
Microsoft Soundscape demonstration, 9 May 2017, London
Vice Chair, All-Party Parliamentary Group on Assistive Technology
Vice Chair, All-Party Parliamentary Group on Financial Technology
Officer, All-Party Parliamentary Group on Fourth Industrial Revolution
Author of report entitled Distributed Ledger Technologies for Public Good: leadership, collaboration and innovation, published November 2017, produced with support from Womble Bond Dickinson LLP and Barclays Bank plc

Lord Levene of Portsoken

Chairman, General Dynamics UK Limited

The Lord Bishop of Oxford

Treasurer, All-Party Parliamentary Group on Artificial Intelligence

Eldest son, Paul Croft, Co-Founder and Creative Director, cluster of games development companies of which the largest is Mediatonic.

Lord Puttnam

Member, Advisory Board of Accenture (Ireland).

Viscount Ridley

Speaking engagement, 20 November 2016, Ciudad de las Ideas 2016

Puebla, Mexico (arranged through Chartwell Speakers' Bureau)

Shareholder in Octopus Protected EIS Fund 7

Baroness Rock

Non-executive Director, Imagination Technologies plc (interest ceased November 2017)

Non-executive Director, Real World Technologies Ltd

Senior Adviser, Instinctif Partners

Attendance, December 2017, Ditchley Foundation Conference, Machine learning and artificial intelligence: how do we make sure technology serves the open society?

Lord St John of Bletso

Non-executive Director, Cognosec Limited

Lord Swinfen

Chairman, The Swinfen Charitable Trust

A full list of Member's interests can be found in the Register of Lords Interests: <http://www.parliament.uk/mps-lords-and-offices/standards-and-interests/register-of-lords-interests/>

Dr Mateja Jamnik (Specialist Adviser)

Reader in Artificial Intelligence, Department of Computer Science and Technology (Computer Laboratory), University of Cambridge

Organisations which have been funding Dr Jamnik's work:

- *University of Cambridge*
- *UK Engineering and Physical Sciences Research Council*
- *The Leverhulme Trust*

Angélica Agredo Montealegre (Specialist Adviser)

No relevant interests declared

APPENDIX 2: LIST OF WITNESSES

Evidence is published online at <http://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/> and available for inspection at the Parliamentary Archives (020 7219 3074).

Evidence received by the Committee is listed below in chronological order of oral evidence session and in alphabetical order. Those witnesses marked with ** gave both oral evidence and written evidence. Those marked with * gave oral evidence and did not submit any written evidence. All other witnesses submitted written evidence only.

Oral evidence in chronological order

- * Professor Nick Bostrom, Director, Future of Humanity Institute [QQ 1–8](#)
- * Professor Dame Wendy Hall, Regius Professor of Computer Science, University of Southampton
- ** Professor Michael Wooldridge, Head of Department and Professor of Computer Science, University of Oxford
- * Rory Cellan-Jones, Technology Correspondent, BBC News [QQ 9–17](#)
- * Sarah O'Connor, Employment Correspondent, Financial Times
- * Andrew Orlowski, Executive Editor, The Register
- * Dr Ing. Konstantinos Karachalios, Managing Director, IEEE-Standards Association [QQ 18–28](#)
- * Professor Alan Winfield, Professor of Robot Ethics, University of the West of England
- * Jeremy Barnett, Barrister, St Pauls Chambers, Leeds and Gough Square Chambers [QQ 29–37](#)
- ** Professor Chris Reed, Professor of Electronic Commerce Law, Queen Mary University of London
- * Professor Karen Yeung, Professor of Law and Director of the Centre for Technology, Ethics, Law and Society at Dickson Poon School of Law, King's College London
- * Dr David Barber, Turing Fellow, The Alan Turing Institute, and Reader in Computational Statistics and Machine Learning, UCL [QQ 38–45](#)
- ** Dr Marko Balabanovic, Chief Technology Officer, Digital Catapult
- ** Dr Timothy Lanfear, Director of EMEA Solution Architecture & Engineering Team, NVIDIA
- * Eileen Burbidge MBE, Partner, Passion Capital [QQ 46–54](#)
- * David Kelnar, Investment Director and Head of Research, MMC Ventures

- * Libby Kinsey, Co-founder, Project Juno
- ** Dr Mercedes Bunz, Senior Lecturer, Communications and Media Research Institute, University of Westminster [QQ 55–64](#)
- ** Elizabeth Denham, UK Information Commissioner, Information Commissioner’s Office
- * Dr Sandra Wachter, Postdoctoral Researcher in Data Ethics and Algorithms, Oxford Internet Institute
- * Olivier Thereaux, Head of Technology, The Open Data Institute [QQ 65–75](#)
- * Javier Ruiz Diaz, Policy Director, The Open Rights Group
- ** Frederike Kalthener, Policy Officer, Privacy International
- ** Dr James Luke, Chief Technology Officer for the Public Sector, IBM [QQ 76–84](#)
- ** Kriti Sharma, Vice President of Artificial Intelligence and Bots, Sage
- * Andrew de Rozairo, Vice President, Customer Innovation and Enterprise Platform, SAP
- * Colin Griffiths, Policy Manager, Citizens Advice [QQ 85–94](#)
- ** Will Hayter, Project Director, Competition and Markets Authority
- ** Olly Buston, CEO and Founder, Future Advocacy [QQ 95–104](#)
- * Professor Dame Henrietta Moore, Director, Institute for Global Prosperity, UCL
- ** Professor Richard Susskind OBE, IT Adviser to the Lord Chief Justice of England and Wales
- * Dr Mark Taylor, Global Strategy & Research Director, Dyson [QQ 105–115](#)
- ** Dr Joseph Reger, Chief Technology Officer EMEIA, Fujitsu
- ** Paul Clarke, Chief Technology Officer, Ocado
- * Dr Julian Huppert, Chair, Independent Review Panel for DeepMind Health [QQ 116–127](#)
- ** Dr Sobia Raza, Head of Science, PHG Foundation
- * Nicola Perrin, Head, Understanding Patient Data, Wellcome Trust
- ** Dr Hugh Harvey, Clinical Artificial Intelligence Researcher and Consultant Radiologist, Guy’s and St Thomas’ NHS Foundation Trust [QQ 128–142](#)
- * Dame Fiona Caldicott, National Data Guardian for Health and Care, Office of the National Data Guardian

- * Professor Martin Severs, Medical Director, NHS Digital
- ** Dr Mark Briers, Strategic Programme Director for Defence and Security, The Alan Turing Institute [QQ 143–152](#)
- * Professor Chris Hankin, Co-Director, Institute for Security Science and Technology, Imperial College London
- * Major Kitty McKendrick, Visiting Fellow, Chatham House [QQ 153–162](#)
- ** Professor Noel Sharkey, Emeritus Professor of Artificial Intelligence and Robotics and Public Engagement, University of Sheffield
- * Mike Stone, Former Chief of Digital and Information Officer, Ministry of Defence
- * Dr Alvin Wilby, Vice President Research, Technical and Innovation, Thales Group
- * Professor Wolfgang Wahlster, CEO and Scientific Director, German Research Centre for Artificial Intelligence (DFKI) [QQ 163–171](#)
- ** Dr Alan Bernstein, President and CEO, Canadian Institute for Advanced Research (CIFAR) [QQ 172–180](#)
- ** Miles Berry, Principal Lecturer, School of Education, University of Roehampton [QQ 181–189](#)
- * Graham Brown-Martin, Author and entrepreneur
- ** Professor Rosemary Luckin, Professor of Learner Centred Design, UCL Institute of Education
- ** The Rt Hon Matt Hancock MP, Minister of State for Digital, Department for Digital, Culture, Media and Sport [QQ 190–200](#)
- ** The Rt Hon the Lord Henley, Parliamentary Under Secretary of State at the Department for Business, Energy and Industrial Strategy
- * Dr Jérôme Pesenti, CEO, BenevolentTech at BenevolentAI [QQ 201–212](#)
- * Professor David Edgerton, Hans Rausing Professor of the History of Science and Technology, and Professor of Modern British History, King’s College London [QQ 213–223](#)
- * Professor Peter McOwan, Vice Principal, Public Engagement and Student Enterprise, Queen Mary University of London
- * Professor Sir David Spiegelhalter, President, Royal Statistical Society, Winton Professor of the Public Understanding of Risk, University of Cambridge and Chair, Winton Centre for Risk and Evidence Communication

Alphabetical list of all witnesses

10x Future Technology	<u>AIC0024</u>
The Academy of Medical Sciences	<u>AIC0210</u>
Accenture UK Limited	<u>AIC0191</u>
Advanced Marine Innovation Technology Subsea Ltd	<u>AIC0038</u>
Agents, Interaction and Complexity (AIC) Group, University of Southampton	<u>AIC0115</u>
AGISI.org	<u>AIC0184</u>
The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB)	<u>AIC0086</u>
The AI Initiative, The Future Society at Harvard Kennedy School	<u>AIC0209</u>
** The Alan Turing Institute (QQ 143–152)	<u>AIC0139</u>
Mr Jaafar Almusaad	<u>AIC0039</u>
Amnesty International	<u>AIC0180</u>
Dr Sally Applin	<u>AIC0172</u>
Arm	<u>AIC0083</u>
Article 19	<u>AIC0129</u>
The Association of Medical Research Charities (AMRC)	<u>AIC0202</u>
Dr Shahar Avin	<u>AIC0150</u>
Baker McKenzie	<u>AIC0111</u>
Balderton Capital (UK) LLP	<u>AIC0232</u>
* Dr David Barber, Turing Fellow, The Alan Turing Institute, and Reader in Computational Statistics and Machine Learning, UCL (QQ 38–45)	
* Jeremy Barnett (QQ 29–37)	
Professor Andrew Basden	<u>AIC0195</u>
BBC	<u>AIC0204</u>
BCS, The Chartered Institute for IT	<u>AIC0049</u>
Dr Simon Beard	<u>AIC0150</u>
** Miles Berry (QQ 181–189)	<u>AIC0247</u>
Big Brother Watch	<u>AIC0154</u>
Big Innovation Centre	<u>AIC0119</u>
Bikal	<u>AIC0052</u>
Dr Richard Billingsley	<u>AIC0201</u>
BioCentre	<u>AIC0169</u>
Bioss International Ltd	<u>AIC0033</u>
Dr Andrew Blick	<u>AIC0064</u>

	Dr Paula Boddington	<u>AIC0067</u>
	Michael Borgeaud	<u>AIC0233</u>
*	Professor Nick Bostrom (QQ 1–8)	
	Braintree	<u>AIC0074</u>
	Mr Philip Bree	<u>AIC0039</u>
	Bristows LLP	<u>AIC0097</u>
	The British Academy	<u>AIC0213</u>
	The British Institute of Facilities Management (BIFM)	<u>AIC0205</u>
	British Standards Institution	<u>AIC0165</u> <u>AIC0231</u>
*	Graham Brown–Martin (QQ 181–189)	
	BSA The Software Alliance	<u>AIC0153</u>
	Dr Aysegul Bugra	<u>AIC0051</u>
	Professor Alan Bundy	<u>AIC0029</u>
**	Dr Mercedes Bunz (QQ 55–64)	<u>AIC0048</u>
*	Eileen Burbidge MBE (QQ 46–54)	
	Eur. Ing. David Burden	<u>AIC0061</u>
	Michael Butterworth	<u>AIC0104</u>
	Cancer Research UK	<u>AIC0219</u>
	Capco	<u>AIC0071</u>
	CBI	<u>AIC0114</u>
*	Rory Cellan-Jones (QQ 9–17)	
	Center for Data Innovation	<u>AIC0043</u>
	Centre for Health Economics University of York	<u>AIC0242</u>
	Centre for Public Impact	<u>AIC0173</u>
	Centre for the Study of Existential Risk	<u>AIC0237</u> <u>AIC0239</u>
	CENTURY Tech	<u>AIC0084</u>
	Matthew Channon	<u>AIC0051</u>
	Charities Aid Foundation	<u>AIC0042</u>
	Mr Thomas Cheney	<u>AIC0098</u>
	Dr Esyin Chew	<u>AIC0166</u>
	Children’s Commissioner for England	<u>AIC0123</u>
**	CIFAR (QQ 172–180)	<u>AIC0136</u>
*	Citizens Advice (QQ 85–94)	
	Donald Clerk	<u>AIC0022</u>
	CognitionX	<u>AIC0170</u>

	Cognitive Finance Group	AIC0010
**	Competition and Markets Authority (QQ 85–94)	AIC0245
	Contact Centre Systems Ltd.	AIC0032
	Cooley (UK) LLP	AIC0217
	Dr Steven Cranfield	AIC0104
	Will Crosthwait	AIC0094
	Darktrace	AIC0243
	Data & Society Research Institute	AIC0221
	Mr Graeme Davis	AIC0054
	Deep Learning Partnership	AIC0027
	Deep Science Ventures	AIC0167
	DeepMind	AIC0234
	Deloitte	AIC0075
	Department of Computer Science, University of Bath	AIC0099
	Department of Computer Science, University of Liverpool	AIC0192
**	Digital Catapult (QQ 38–45)	AIC0175
	Doteveryone	AIC0148
	Reverend Dr Lyndon Drake	AIC0108
*	Dyson (QQ 105–115)	
	Richard Ebley	AIC0026
	The Economic Singularity Supper Club	AIC0058
*	Professor David Edgerton (QQ 213–223)	
	Professor Lilian Edwards	AIC0161
	Electronic Frontier Foundation	AIC0199
	Dr Julian Estevez	AIC0021
	euRobotics Topics Group on ‘Ethical, Legal and Socio-economic issues’	AIC0189
	Faethm Pty Ltd	AIC0141
	Family Law Partners	AIC0089
	Dr Jerry Fishenden	AIC0028
	Professor Robert Fisher	AIC0029
	Dr Malcolm Fisk	AIC0012
	Five AI Ltd	AIC0128
	Foundation for Responsible Robotics	AIC0188
	Professor John Fox	AIC0076
	Laurence Freeman	AIC0147

**	Fujitsu (QQ 105–115)	AIC0120
**	Future Advocacy (QQ 95–104)	AIC0121
	Future Intelligence	AIC0216
	Future of Humanity Institute	AIC0103
	Dr Samantha Gallivan	AIC0185
*	German Research Centre for Artificial Intelligence (DFKI) (QQ 163–171)	
	Joanna Goodman	AIC0104
	Google	AIC0225
	Government of Canada	AIC0222
	Government of China	AIC0145
	Government of Japan	AIC0224
	Government of the Republic of Korea	AIC0228
	Dr Paul Graham	AIC0088
	Guide Dogs	AIC0040
	Dr Ozlem Gurses	AIC0051
*	Professor Dame Wendy Hall (QQ 1–8)	
*	Professor Chris Hankin (QQ 143–152)	
	Baroness Harding of Winscombe	AIC0072
*	Dr Hugh Harvey (QQ 128–142)	
**	HM Government (QQ 190–200)	AIC0229
	Fabia Howard-Smith	AIC0147
	The Human Rights, Big Data and Technology Project	AIC0196
*	Dr Julian Huppert (QQ 116–127)	
	Dr Catrin Fflûr Huws	AIC0008
**	IBM (QQ 76–84)	AIC0160
	IEEE European Public Policy Initiative Working Group on ICT	AIC0106
	The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems	AIC0100
*	IEEE-Standards Association (QQ 18–28)	
	Imperial College London	AIC0214
**	Information Commissioner’s Office (QQ 55–64)	AIC0132
	Information Systems Audit and Control Association (ISACA) London Chapter	AIC0193
	Information Technology Industry Council (ITI)	AIC0176
	Innovate UK	AIC0220

The Institute of Chartered Accountants in England and Wales	<u>AIC0041</u>
Institute of Mathematics and its Applications	<u>AIC0107</u>
International Associates	<u>AIC0003</u>
Dr Maria Ioannidou	<u>AIC0082</u>
Brian Joyce	<u>AIC0179</u>
Dr Paresh Kathrani	<u>AIC0104</u>
Kemp Little LLP	<u>AIC0133</u>
Professor Simon King	<u>AIC0029</u>
Dr Ben Kirman	<u>AIC0127</u>
The Knowledge, Skills and Experience Foundation	<u>AIC0044</u>
Dr Ansgar Koene	<u>AIC0208</u>
Dr Antonios Kouroutakis	<u>AIC0051</u>
KPMG LLP	<u>AIC0211</u>
Martina Kunz	<u>AIC0150</u>
Maciej Kuziemski	<u>AIC0197</u>
Professor Marta Kwiatkowska	<u>AIC0190</u>
Dale Lane	<u>AIC0059</u>
Law and Innovation Research Group and the Legal Teaching Research Group from The Fundação Getúlio Vargas School of Law, São Paulo, Brazil	<u>AIC0177</u>
Dr David Lawrence	<u>AIC0036</u>
The Law Society of England and Wales	<u>AIC0152</u>
James Lawson	<u>AIC0073</u>
Professor Shaun Lawson	<u>AIC0127</u>
Professor Mark Lee	<u>AIC0093</u>
Leverhulme Centre for the Future of Intelligence	<u>AIC0182</u>
	<u>AIC0236</u>
	<u>AIC0237</u>
	<u>AIC0238</u>
	<u>AIC0239</u>
LexisNexis UK	<u>AIC0164</u>
Liberty	<u>AIC0181</u>
Chrissie Lightfoot	<u>AIC0104</u>
Dr Conor Linehan	<u>AIC0127</u>
** Professor Rosemary Luckin (QQ 181–189)	<u>AIC0246</u>
Dr Mike Lynch	<u>AIC0005</u>
	<u>AIC0230</u>

Ms Nika Mahnič	AIC0200
The Market Research Society	AIC0130
Professor James Marshall	AIC0088
Dr Neil McBride	AIC0047
* Major Kitty McKendrick (QQ 153–162)	
Mr John McNamara	AIC0081
Sabine McNeill	AIC0009
* Professor Peter McOwan (QQ 213–223)	
Professor Andrew McStay	AIC0015
medConfidential	AIC0063 AIC0244
The Medicines and Healthcare products Regulatory Agency (MHRA)	AIC0134
Microsoft	AIC0149
* MMC Ventures (QQ 46–54)	
* Professor Dame Henrietta Moore (QQ 95–104)	
Dr Zdenek Moravcik	AIC0019
Dr Ian Morgan	AIC0179
Dr Sarah Morley	AIC0036
** National Data Guardian for Health and Care (QQ 128–142)	AIC0143
Professor John Naughton	AIC0144
NCC Group plc	AIC0240
Dr Jean-Christophe Nebel	AIC0102
Hadley Newman	AIC0155 AIC0156
* NHS Digital (QQ 128–142)	
Nominet	AIC0131
Norton Rose Fulbright LLP	AIC0079
Professor Thomas Nowotny	AIC0088
** NVIDIA (QQ 38–45)	AIC0212
** Ocado Group plc (QQ 105–115)	AIC0050
Mr Jeremy O'Connor	AIC0034
* Sarah O'Connor (QQ 9–17)	
Dr Dan O'Hara	AIC0127
Dr Seán Ó hÉigearthaigh	AIC0150
Dr James O'Shea	AIC0226
Alex Olson	AIC0002

	Onfido	AIC0163
	Online Dating Association	AIC0110
*	Open Data Institute (QQ 65–75)	
*	Open Rights Group (QQ 65–75)	
	ORBIT The Observatory for Responsible Research and Innovation in ICT	AIC0109
	Ordnance Survey	AIC0090
*	Andrew Orłowski (QQ 9–17)	
	Marion Oswald	AIC0068
	Professor Maja Pantic	AIC0215
	Dr Andrew Pardoe	AIC0020
	Joshua Parikh	AIC0031
	Jonathan Penn	AIC0198
*	Dr Jérôme Pesenti (QQ 201–212)	
**	PHG Foundation (QQ 116–127)	AIC0092
	Dr Andrew Philippides	AIC0088
	Toby Phillips	AIC0197
	Professor Barbara Pierscionek	AIC0046
	Professor John Preston	AIC0014
	PricewaterhouseCoopers LLP (PwC)	AIC0162
**	Privacy International (QQ 65–75)	AIC0207
*	Project Juno (QQ 46–54)	
	Raymond Williams Foundation	AIC0122
**	Professor Chris Reed (QQ 29–37)	AIC0055
	Research Councils UK	AIC0142
	Research into Employment, Empowerment and Futures Centre (REEF), The Open University	AIC0124
	Professor Kathleen Richardson	AIC0200
	Professor David Robertson	AIC0029
	Mrs Violet Rook	AIC0151
	Dr Michael Rovatsos	AIC0029
	Royal Academy of Engineering	AIC0140
	The Royal College of Radiologists	AIC0146
	The Royal Society	AIC0168
	The Royal Statistical Society	AIC0218
	The RSA	AIC0157
	Dr John Rumbold	AIC0046

	SafeToNet	AIC0087
**	Sage (QQ 76–84)	AIC0159
*	SAP (QQ 76–84)	
	Professor Maggi Savin-Baden	AIC0061
	SCAMPI Research Consortium, City, University of London	AIC0060
	Dr Valentina Rita Scotti	AIC0051
	Dr Huma Shah	AIC0066
**	Professor Noel Sharkey (QQ 153–162)	AIC0248
	Simul Systems Ltd	AIC0016
	Jonathan Sinclair	AIC0023
		AIC0035
	SiteFocus Incorporated	AIC0187
	Dr Will Slocombe	AIC0056
*	Professor Sir David Spiegelhalter (QQ 213–223)	
	Dr Chris Steed	AIC0017
*	Mike Stone (QQ 153–162)	
**	Professor Richard Susskind OBE (QQ 95–104)	AIC0194
	Professor Austin Tate	AIC0029
	techUK	AIC0203
*	Thales Group (QQ 153–162)	
	Thames Valley Police	AIC0125
	Thomson Reuters	AIC0223
	Touch Surgery	AIC0070
	Transport Systems Catapult	AIC0158
	Richard Tromans	AIC0227
	UCL Knowledge Lab	AIC0105
	UK Computing Research Committee	AIC0030
	The Association for UK Interactive Entertainment (Ukie)	AIC0116
*	Understanding Patient Data, Wellcome Trust (QQ 116–127)	
	Dr Ozlem Ulgen	AIC0112
	University College London (UCL)	AIC0135
	Sheena Urwin	AIC0068
	Michael Veale	AIC0065
	Professor Chris Voss	AIC0118
*	Dr Sandra Wachter (QQ 55–64)	

Dr Toby Walsh	<u>AIC0078</u>
Andrew Ware	<u>AIC0150</u>
Professor Kevin Warwick	<u>AIC0066</u>
Warwick Business School University of Warwick	<u>AIC0117</u>
Weightmans LLP	<u>AIC0080</u>
Wellcome Trust	<u>AIC0202</u>
Vishal Wilde	<u>AIC0004</u>
Professor Chris Williams	<u>AIC0029</u>
Professor Rebecca Williams	<u>AIC0206</u>
* Professor Alan Winfield (QQ 18–28)	
** Professor Michael Wooldridge (QQ 1–8)	<u>AIC0174</u>
Workday Inc.	<u>AIC0183</u>
* Professor Karen Yeung (QQ 29–37)	
Young Enterprise	<u>AIC0091</u>
Dr Jianhan Zhu	<u>AIC0045</u>
Diego Zuluaga	<u>AIC0235</u>

APPENDIX 3: CALL FOR EVIDENCE

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.

This is a public call for written evidence to be submitted to the Committee. The deadline is 6 September 2017.

We are looking to hear from as many people and organisations as possible—if you think someone you know would have views to contribute, please do pass this on to them.

When preparing your response, please bear in mind that short, concise submissions are preferred and responses must not be any longer than six sides of A4. We do not expect you to address every question below. How to submit evidence is set out later in this document but if you require any questions or adjustments to enable you to respond, please contact the staff of the Committee on the details provided.

We are looking for pragmatic solutions to the issues presented by artificial intelligence, so please provide practical examples where possible. Finally, we would be interested to know how you have defined artificial intelligence in your response.

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?
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?

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

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?

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.

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?

APPENDIX 4: HISTORIC GOVERNMENT POLICY ON ARTIFICIAL INTELLIGENCE IN THE UNITED KINGDOM

The Committee appointed Angelica Agredo Montealegre, a PhD student at King's College London, as a Specialist Adviser to conduct research into historic Government policy on artificial intelligence. The following is the result of that work.

This note presents a long-term perspective on artificial intelligence (AI) research and development (R&D) in the United Kingdom in the last 40 years. In particular, it outlines the results and implementation problems of Government-supported AI R&D projects.

The note is divided into three main parts. The first part provides an overview and shows that the way AI has been understood in the past 40 years has varied considerably, and that it has usually been considered as an aspect of information technology (IT) policy, rather than an entirely distinctive field. The second part presents the Alvey programme (1983–1987)—the only large-scale national project of the sort during this period—and outlines its objectives, achievements and execution problems. The final part draws a comparison between the Alvey programme and its counterparts in Japan, Europe and the USA.

Overview

The potential of AI generated great enthusiasm and high expectations in the 1950s, leading to the formation of a number of major AI research centres in the UK at the universities of Edinburgh, Sussex, Essex and Cambridge in the 1960s. But by the 1970s this enthusiasm had begun to wane, as promises went unfulfilled, both in the UK and in the USA, the other major global centre of AI, and members of the research community became embroiled in fierce disputes regarding the nature and aims of AI research. This discord, along with the broader backdrop of disappointment, prompted the Science Research Council to commission an inquiry into the state of the field. The resulting report, produced by Professor Sir James Lighthill in 1973, made clear his pessimism about the potential outcomes of basic research in AI in the near future.⁵⁸⁶ The Lighthill Report is now mostly known within the AI community for causing a reduction in support for AI research in the UK, a period also known as the first 'AI winter'.

The 1980s saw the creation of the Alvey programme (1983–1987), the first large-scale R&D project involving AI in Britain. Before this programme there were no large national projects featuring AI; instead the Government funded AI through the Science Research Council at universities such as Edinburgh and Cambridge. The launch of the Alvey programme was a response to the creation of the Japanese Fifth Generation Computer programme in 1982. After its success in the electronic consumer goods and automobile industries, Japan announced its attempt to create a new computer with the capacity to solve problems on its own. In order to facilitate its use, the creation of this computer was accompanied by developments in the human-machine interface (in the same way that virtual assistants such as Siri and Alexa are attempting to do today). In this context, the Alvey programme was intended to establish Britain as a key player in the IT sector worldwide.

⁵⁸⁶ Science Research Council, *Artificial Intelligence: A paper symposium* (1973): http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm [accessed 5 February 2018]

However, by the late 1980s, the UK's IT sector had built up a considerable trade deficit, and by the early 1990s it was considered unlikely that the Alvey programme would lead to any substantive commercial returns.⁵⁸⁷ The evaluation commissioned by the Department of Trade and Industry after the programme stated that the Alvey programme's focus on pre-competitive research was one of the main factors that hindered its contribution to the enhancement of the UK's competitiveness in the IT market.⁵⁸⁸ The Government subsequently rejected proposals to create a follow-up project and the aftermath of the Alvey programme is sometimes referred to as the second 'AI winter', which lasted until the early 1990s. This coincided with a general loss of enthusiasm for AI in the US.⁵⁸⁹

In the UK in the 1990s, support for AI research stopped being systematic, and there were no attempts to co-ordinate research at a national scale. Instead, the UK's participation in the European research project ESPRIT was intensified.⁵⁹⁰ Even though no national large-scale programmes were created in the UK, AI research continued, both in industry and in the same universities which had led the way in the early 1970s—albeit at a reduced scale. More recently, emphasis shifted towards privately funded research. As such, the Alvey programme represented the first and last major government-funded AI project in the UK.

It should be noted that a lack of clarity in terms of definitions and objectives seems to have plagued the field right back to its origins in the 1950s. This makes tracing the evolution of the AI field in the UK a difficult task. It is usually unavoidable to refer to IT in general and, even then, information is scant. For instance, the governments of the day appear not to have collected comprehensive, systematic data related to the amount of funds spent on IT R&D, let alone specifically on AI. As such, the best source of information that is available about AI development in late twentieth-century Britain relate to the Alvey programme. Indeed, after the Alvey programme, AI R&D policy was spread between different funding councils and stopped being systematic, making it very difficult to trace.⁵⁹¹

587 Science Policy Research Unit, University of Sussex and Programme of Policy Research in Engineering, Science and Technology, University of Manchester, *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, (Norwich, United Kingdom: Her Majesty's Stationery Office, 1991), p iv

588 The term 'pre-competitive research' was intended to delineate so-called enabling technologies, which on their own did not have viable commercial applications, but were considered necessary for the subsequent development of commercially competitive products and systems by private companies. Brian Oakley and Kenneth Owen, *Alvey: Britain's Strategic Computing Initiative* (Cambridge, Massachusetts and London: The MIT Press, 1989), p 3

589 *Thinking Machines: The Quest for Artificial Intelligence—and where it's taking us next*, p 3

590 The European Strategic Program on Research in Information Technology (ESPRIT) was a series of integrated programmes of information technology research and development projects and industrial technology transfer measures which ran from 1983 to 1998.

591 The National Archives, Records of the Prime Minister's Office: Correspondence and Papers, 1979–1997, PREM 19/2116: 'EDUCATION. New blood for research and information technology: follow-up to the Alvey Report on Advanced Information Technology; international collaboration; the EUREKA programme', (1982–1987): <http://discovery.nationalarchives.gov.uk/details/r/C16204995> [accessed 23 March 2018]

Defining artificial intelligence and its aims

The definition of AI has changed over time, and variations on the meaning can be found within the same period—as AI pioneer John McCarthy once put it, “as soon as it works, no one calls it AI any more”. This lack of precision has historically generated confusion amongst policy makers and members of the IT community—as it still does today. Indeed, in the past, there was no consensus about what the goals of AI research should be, or how best to achieve them. This is a crucial issue because the Government, researchers, suppliers and users have had different ideas of what the role of AI is, as well as its potential, contributing to disappointment from different parties when these expectations have not been met.

In the 1970s, there was no consensus amongst researchers about the definition of AI, its objectives, or its economic potential in either the short or the long term. In his report, Professor Sir James Lighthill attempted a description of the AI field that generated various criticisms. Professor Lighthill divided the field into three categories: advanced automation, computer-based central nervous system research, and “robot-building”.⁵⁹² The objectives of these distinct areas were, respectively:

- the creation of machines to replace human beings for specific tasks;
- the emulation of functions of the brain for research purposes in neurobiology; and
- the creation of automatic devices mimicking a range of human functions without seeking to replace human beings “in any useful sphere of human activity”.⁵⁹³

Professor Lighthill stated that research in AI would probably not yield significant achievements in the following 25 years, and that the creation of an intelligent, general-purpose system was a goal that would not be fulfilled in the twentieth-century—if ever.⁵⁹⁴ However, he did point out that the chances of success in AI would increase if research was integrated with the field of application.⁵⁹⁵ Professor Lighthill’s categorisation and conclusions were contested by leading figures of the field in the UK, such as Professor Stuart Sutherland (founder and head of the University of Sussex’s Laboratory of Experimental Biology), and Professor Donald Michie (director of the University of Edinburgh’s Department of Machine Intelligence and Perception). They suggested replacing the category of “robot-building” with basic research.⁵⁹⁶ They were also more optimistic about the field’s achievements and promises, and stated that a large investment in basic AI was justified if Britain’s AI field was to be competitive worldwide.⁵⁹⁷

In the 1980s, the UK AI research community coined its own term for AI: Intelligent Knowledge Based Systems (IKBS). IKBS described “systems which combine hardware and software in order to achieve the goal of using inference to apply knowledge to perform a task”.⁵⁹⁸ However, ‘AI’ was still used, especially when referring to ‘Expert Systems’—computer systems that attempt to emulate the decision-making of a human by using a set of facts and rules that had been

592 *Artificial Intelligence: A paper symposium*, p 3

593 *Artificial Intelligence: A paper symposium*, pp 5–8

594 *Artificial Intelligence: A paper symposium*, p 15

595 *Artificial Intelligence: A paper symposium*, p 18

596 *Artificial Intelligence: A paper symposium*, p 20

597 Professor Sutherland also recommended improving the connections with the USA given the highly developed state of its AI field. He suggested doing this by facilitating and encouraging the movement of students and researchers between the two countries. *Artificial Intelligence: A paper symposium*, p 26

598 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p 15

previously programmed into them. Even within the category of ‘Expert Systems’ there was no consensus amongst researchers about the aims of the field, which caused delays in the acceptance of research grants (as was the case, for instance, at the Medical Research Council).⁵⁹⁹

In terms of wider R&D policy, the Government usually considered AI as an aspect of IT policy, rather than an entirely distinctive field. Indeed, it should be noted that AI (or IKBS) was only one aspect of the Alvey programme, which meant that the benefits expected from the project were expected to be advancements in the IT sector, not just AI.

The Alvey programme

The Alvey programme was a five-year collaborative R&D programme in IT which began in 1983. It was funded by the Department of Trade and Industry (DTI), the Science Engineering Research Council (SERC) and the Ministry of Defence (MoD). In total the programme cost £350 million (approximately £1 billion today), of which £200 million came from the Government and the remainder from industry.

Not all funding was new. For instance, the Very Large-Scale Integration (VLSI) sub-programme was based on a similar programme already planned by the MoD.⁶⁰⁰ Over 300 sub-projects were initiated, nearly 200 of them involving both industrial and academic research teams. The remainder, about 8% of the project budget, were smaller academic-only sub-projects in which companies took an interest. The research strategy was determined by a Directorate, staffed partly by industrial and academic secondees, which oversaw the formulation and implementation of the programme strategy.

Objectives and achievements

The programme had three main categories of objective:

- Technological: these were related to the development of pre-competitive, advanced IT. Research was focused on so-called enabling technologies, and not on producing particular products.
- Structural: these were related to the consolidation of the IT sector in the UK. At the beginning of the 1980s it was considered that the sector was weak and fragmented, and the Alvey programme was intended to broaden the UK’s IT R&D base and consolidate it.
- Strategic: these were intended to preserve the UK’s IT capability and improve its economic competitiveness and performance in relation to other countries, particularly the US and Japan. Although most of the work in the Alvey programme was not intended to lead directly to production, the programme was expected to improve the position of British firms to realise the commercial potential of R&D.⁶⁰¹

599 For instance, in the 1980s, the Medical Research Council received a number of research grant applications that included ‘Expert Systems’ and referred to them as the application of ‘AI’ to medicine. The National Archives, Medical Research Council: Registered Files, Scientific Matters (S Series), FD 23/2286: ‘Artificial Intelligence Advisory Group: grant applications considered by the Group; notes and correspondence’ (1986): <http://discovery.nationalarchives.gov.uk/details/r/C2516889> [accessed 23 March 2018]

600 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p i

601 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p iii

Upon the conclusion of the Alvey programme, participants deemed the technological objectives to have been met. They believed that the programme had correctly identified and supported certain enabling technologies, which remained critical at the beginning of the 1990s. The programme as a whole was primarily focused on producing software rather than hardware.⁶⁰² The achievements in IKBS were considered to be lower than in other areas of the programme.⁶⁰³

Following the end of the programme, the DTI commissioned an evaluation of its achievements to be carried out by the Science Policy Research Unit at the University of Sussex, and the Programme of Policy Research in Engineering, Science and Technology at the University of Manchester. These evaluators deemed the structural objectives to have been successfully met. The number of researchers increased, with over 5,000 people involved in the programme.⁶⁰⁴ Moreover, the quality and extent of communications and links between the members of the growing IT community were improved. However, no new research centres of any significance emerged after the programme. The Alvey programme probably benefitted from the existence of important research centres, such as Cambridge, but did not appear to have made significant changes to the size or distribution of research activities in the UK.⁶⁰⁵

Despite these apparent successes, the strategic objectives were not achieved. By 1980, the UK IT sector had a trade deficit of £300 million, and while the directors of the programme projected it would reach £1 billion by 1990, in reality it was surpassed as early as 1984.⁶⁰⁶ Although it was expected that in the long-term the programme's work could lead to commercial returns, by the early 1990s it was considered that these expectations were unlikely to be met.⁶⁰⁷ Although it was never made explicit what was meant by 'long-term', the time ranges given were usually of 10 years or more.

According to the evaluators, the obstacles for the commercial exploitation of the Alvey programme's work were similar to the ones present in any collaborative R&D project at this time. Firms often changed strategy, which affected their partners' expected exploitation channels. Moreover, inadequate management of the connections between R&D and production was a common issue, which was further exacerbated by capital shortages. Collectively, these all worked to hamstring attempts to move from research to production.⁶⁰⁸

602 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p 141

603 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p 144

604 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p iii

605 The Cambridge advanced IT 'phenomenon' was already happening beforehand. Brian Oakley and Kenneth Owen, *Alvey: Britain's Strategic Computing Initiative* (Cambridge, Mass and London: The MIT Press, 1989), p 111

606 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p 5

607 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p iv

608 *Ibid.*

Implementation problems of the Alvey programme

During implementation the programme encountered a variety of problems. The director of the programme from 1983 to 1987, Brian Oakley, identified the difficulty of financing the academic part of the programme as the most challenging issue he had to deal with. In his view, academic funds were over-allocated, which resulted in cash flow problems.⁶⁰⁹

Moreover, Mr Oakley pointed out that many research assistants working on the Alvey programme were not from the UK, and most subsequently returned to their country of origin at the end of their contract.⁶¹⁰ He saw this as a problem because skilled staff trained by the programme were not subsequently benefiting the UK. Mr Oakley suggested that several research assistants had probably left academia to join industry because of higher salaries, but he lamented the lack of data collected on the subject. On a more positive note, he observed that some assistants had helped organise IKBS schemes at the Glasgow Turing Institute and at Imperial College, which had provided opportunities to graduate students to work with AI experts for six months.⁶¹¹

A related problem was the lack of qualified staff in both industry and academia. The scarcity of staff was frequently mentioned as a persistent issue throughout the programme and the evaluators in 1991 argued that education and training activities complementary to the Alvey programme should have been put in place to alleviate the situation.⁶¹² Neither the directors nor the evaluators of the programme discussed the reason for this scarcity of qualified staff. The evaluators of the programme suggested that the lack of qualified staff was to be expected given the novelty of the research field.⁶¹³

In many ways, this shortage was surprising, as Britain had led the world in electronic computing during, and after, the Second World War, and it appeared that staff shortages only became prevalent during the 1970s. The historian of technology Dr Marie Hicks has argued that the British government's persistent computer labour problems were in great part a result of the continued neglect of women's labour from the 1960s onwards.⁶¹⁴ Women had dominated the computing workforce in the initial post-war decades, when many positions were considered essentially clerical in nature, and therefore had less status attached to them. But by the 1970s, as computing jobs shifted from machine operations towards higher level management and strategic roles, women were increasingly passed over, even as vacancies at the top continued to grow.⁶¹⁵ The female workforce gradually concentrated in lower-paid and part-time employment and by the mid-1980s, 45% of all women employed in computing were hired only part time.⁶¹⁶

609 Brian Oakley and Kenneth Owen, *Alvey: Britain's Strategic Computing Initiative* (Cambridge, Mass and London: The MIT Press, 1989), p 104

610 *Alvey: Britain's Strategic Computing Initiative*, p 106

611 This does not refer to the present-day Alan Turing Institute based in London, but to the Turing Institute based in Glasgow between 1983 and 1994. *Alvey: Britain's Strategic Computing Initiative*, p 117

612 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p vii

613 *Ibid.*

614 Marie Hicks, *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing* (Cambridge, MA: The MIT Press, 2017), p 208

615 *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*, p 207

616 *Programmed Inequality: How Britain Discarded Women Technologists and Lost Its Edge in Computing*, pp 214–216

Despite Mr Oakley's claims that the Alvey programme contributed to the expansion of trained staff, by the 1990s the shortage of skilled staff remained a problem.⁶¹⁷ If indeed it was the case that a large number of researchers in the Alvey programme were from overseas, this raises further questions as to why the UK was unwilling or unable to retain them after they completed their contracts.

A further source of problems for the Alvey programme was related to the support it attempted to provide for the British computing industry. Although the programme was meant to contribute to the UK's competitiveness in the sector through pre-competitive research, measures were taken to advance this goal more directly. This assistance took two forms: capital purchases and small firm participation. At the beginning of the programme, the Alvey Directorate decided to make bulk purchases of British computing equipment. Buying the same equipment for the participants was meant to facilitate communication between different research projects.

However, not everyone involved with the Alvey programme agreed with the decision to buy British equipment, with some voicing concerns about quality. Even Mr Oakley later admitted that these purchasing decisions were, in hindsight, "an expensive mistake".⁶¹⁸ Aaron Sloman, Professor of AI and Cognitive Science at the University of Sussex, cited it as an error that should be avoided in future programmes, in his letter to Sir Austin Bide in 1986: "[this mistake was] forcing people to use totally unsuitable hardware and software just because it is British, thereby holding up significant research and diverting precious highly skilled manpower from advanced research to low-level software development".⁶¹⁹

It was also unclear afterwards what had been gained from involving small firms in the project. Of about 113 firms that participated in the programme, at least 50 were SMEs (according to the European Commission's definition of SMEs as having less than 400 employees).⁶²⁰ Several of these small firms played invaluable parts in some projects and acquired technology with a relatively small investment. They probably also benefited from the programme's publicity.⁶²¹ However, many small firms struggled to find the staff they needed, as well as secure their share of the research funds. The overhead of working on a co-operative project was such that many small firms decided to form partnerships with larger ones and effectively act as sub-contractors.⁶²² The evaluators of the programme indicated that to some extent the overheads were fixed costs and thus they often made collaborative research less attractive for smaller projects and smaller firms.⁶²³

Communication between the participants was another challenging area. The programme was not organised around one research centre (unlike the Japanese and American programmes—see below). Instead it had been intentionally decentralised, as distances within the UK were relatively small, academics had teaching responsibilities in their universities, and the relocation of the researchers' families presented difficulties.⁶²⁴ Moreover, it was considered that a single site would exacerbate the problems of technology transfer back into the firms for exploitation.⁶²⁵

617 *Alvey: Britain's Strategic Computing Initiative*, p 117

618 *Alvey: Britain's Strategic Computing Initiative*, p 109

619 A. Sloman, Letter to Sir Austin Bide, (undated) 1986, cited in: *Alvey: Britain's Strategic Computing Initiative*, p 158

620 *Alvey: Britain's Strategic Computing Initiative*, p 111

621 *Ibid.*

622 *Ibid.*

623 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p iii

624 *Alvey: Britain's Strategic Computing Initiative*, p 113

625 *Ibid.*

The Alvey programme compared

The Alvey programme was launched in a context of global enthusiasm for large-scale advanced IT research projects. The Japanese Fifth Generation Computer programme, started in 1982, was a ten-year initiative that had the objective of creating a new kind of computer. The term ‘Fifth Generation’ was coined by the Japanese to describe what they thought was going to be the new wave of computer developments. The new computer would use VLSI circuits and new software languages, it would be structured to process information in parallel rather than in sequence, it would exploit a new human-machine interface using speech and image devices and, most importantly, it would be geared towards problem-solving using developments in AI to manage concepts rather than numbers.⁶²⁶

At the end of the ten-year period, the programme had spent over 50 million JPY (the equivalent of £416 million today). This initiative was seen by the European and the American IT sectors as an attempt from the Japanese government to place the country in an advantageous position in the area of computing—as they had done with electronic consumer goods and, to an extent, with the motor vehicle industry.⁶²⁷

The USA saw the Japanese programme as a threat to their global supremacy over the computer and informatics industries.⁶²⁸ Two programmes were subsequently created: the Strategic Computing Initiative (SCI) and the Microelectronics and Computer Technology Corporation (MCC). The SCI was a ten-year programme intended to develop advanced computer hardware and AI. It cost \$1 billion USD (the equivalent of £1.85 billion today), provided by the Defense Advanced Research Projects Agency (DARPA). The SCI was conceived as an integrated programme, which meant that different subsystems were created, all working towards the same goal: creating machine intelligence. The MCC was the first American computer industry R&D consortia. The initial budget for MCC’s activities was between \$50 and \$100 million USD per year (the equivalent of £92 to £184 million today), depending on the number of participants and projects.⁶²⁹

Europe’s response to the Japanese programme was the European Strategic Programme for Research and Development in Information Technology (ESPRIT). The main objective of this initiative was to promote collaboration between European countries in IT R&D. Indeed, every project had to bring together companies and research institutions from at least two EEC countries.⁶³⁰ Moreover, a significant part of the programme was devoted to increase the interaction between users and developers, disseminate results widely, and boost product and process adoption in the market. In the early 1990s, 20% of the overall funding was dedicated to the integration between R&D and take-up.⁶³¹ The programme of ESPRIT was not fixed from the start: it was adapted every year after extensive consultation with both suppliers and users, in order to reflect the industry’s changing priorities.

626 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p 5

627 *Ibid.*

628 Japan had been accused by members of the American IT industry of profiting from other countries’ innovations, which they considered unfair. Copying (as opposed to inventing and innovating) was not considered a threat. Andrew Pollack, ‘Fifth Generation’ became Japan’s Lost Generation’, *New York Times* (5 June 1992): <https://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html> [accessed 23 March 2018]

629 Michael Frontain, ‘Microelectronics and Computer Tech Corporation (MCC)’, *Handbook of Texas Online* (15 June 2010): <https://tshaonline.org/handbook/online/articles/dnm01> [accessed 6 September 2017]

630 European Commission, Community Research and Development Information Service (CORDIS): www.cordis.europa.eu/esprit/src/intro.htm [accessed 6 September 2017]

631 *Ibid.*

There are a number of observations that emerge from the comparison between these projects. First, the Alvey programme was the shortest programme. In the United States, MCC was dissolved in 2000 and the SCI lasted for ten years, like the Japanese Fifth Generation initiative. At its start, ESPRIT had goals that extended over ten years, but subsequent programmes were created and it was only in 1999 that ESPRIT 4 was replaced by the Information Society Technologies (IST) programme.

The Alvey Directorate assumed from early on that there would be a follow-up programme, and in some instances built programmes that were only viable if they lasted ten years.⁶³² The Government and industry, however, saw the Alvey programme as a one-off five-year programme intended to stimulate the community and make the UK an effective international partner or competitor.⁶³³ It is not clear why there was such a disparity of views. In 1988 the Government rejected the proposal of a follow-up programme and instead decided to put more emphasis on the UK's participation in ESPRIT. Indeed the UK began to participate in ESPRIT, on a relatively marginal basis, during the 1980s, but the Alvey programme had absorbed most of the attention and resources during this period. The misunderstanding in terms of the expectations and timeframe of the programme seems to have contributed to disappointment from both parties, as the Directorate was expecting more time to fulfil its goals, while the Government expected results that the programme was never designed to achieve.

The evaluators of the Alvey programme noted that supporting pre-competitive research was a necessary but limited aspect in enhancing the competitiveness and performance of the UK's IT industry.⁶³⁴ Therefore, they suggested that if improving competitiveness was the objective, governmental and private initiatives had to complement pre-competitive R&D.⁶³⁵ These complementary measures should include:

“more concerted efforts to involve IT users with the scope of R&D programmes, thus stimulating greater user awareness and alerting IT producers at an early stage to the needs of users;

greater effort should be made within firms to formulate technology strategies to facilitate the exploitation of R&D. In some cases it might be appropriate for Government to encourage the process;

a serious re-evaluation of mechanisms to cope with the need for ‘patient’ capital in the further development and exploitation of the enhanced R&D base created by programmes such as Alvey.”⁶³⁶

ESPRIT, and to a lesser extent the American programmes, gave significant support to the interaction between users and developers, the dissemination of results, and the promotion of production and adoption of IT products in the market.⁶³⁷ Drawing from this experience, proposals for an ‘After-Alvey’ follow-

632 John Fairclough (Government's Chief Scientific Adviser) cited in: *Alvey: Britain's Strategic Computing Initiative*, p 259

633 *Alvey: Britain's Strategic Computing Initiative*, p 260

634 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p vii

635 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p viii

636 *Ibid.*

637 The American programmes also granted considerable attention to users' needs.

up programme suggested a heavier focus on increasing the application of IT and on enabling a closer involvement between suppliers, researchers and users.⁶³⁸ However, with the shift towards European research projects in the late 1980s, this programme was never approved.

Another important lesson that the evaluators and directors of the Alvey programme drew from the comparison between the implementation of the programme and that of ESPRIT was related to the terms of collaboration between members. At the beginning of the programme, the Alvey Directorate had assumed that firms would have had prior experience of drawing up collaborative agreements; since this was not the case, the start of many projects was delayed.⁶³⁹ The evaluators thus suggested that future programmes should emulate ESPRIT's standard agreements embodied in the contracts, with intellectual property rights held by all participants.⁶⁴⁰

Finally, a brief comparison of the results of these programmes can be fruitful. The fact that ESPRIT facilitated communication between users and developers granted the programme a great deal of flexibility. The annual revision of ESPRIT's programme, along with the establishment of industrial advisory panels (which made recommendations from the perspective of IT users), kept the programme in line with the demands of the IT sector.⁶⁴¹ Moreover, additional support was provided to firms, and especially SMEs, that wanted to incorporate novel electronics (to them) into existing product lines.⁶⁴² The diffusion service, PROSOMA, also served an important role in rendering visible the way in which companies within the IT industry had used their participation in ESPRIT to improve their competitiveness in world markets, as well as both publicising and facilitating the use of ESPRIT results by different agents.⁶⁴³

The areas in which ESPRIT results were applied were greatly varied, and included:

- transport;
- medicine;
- energy;
- electronics;
- manufacturing;
- engineering;
- telecommunications; and
- education.

Although the SCI did not provide such an open exchange mechanism, DARPA was concerned with investing in technologies which showed the most promise. Therefore, in 1987 its new director Dr Jack Schwartz decided to abandon the

638 *Alvey: Britain's Strategic Computing Initiative*, p 229

639 *Alvey: Britain's Strategic Computing Initiative*, pp 109–110

640 *Evaluation of the Alvey Programme for Advanced Information Technology: A Report by Science Policy Research Unit, University of Sussex, and Programme of Policy Research in Engineering, Science and Technology, University of Manchester*, p xi

641 European Commission, Community Research and Development Information Service (CORDIS), 'Archives': www.cordis.europa.eu/esprit/src/intro.htm [accessed 8 March 2018]

642 This service was called FUSE (First Users).

643 European Commission, Community Research and Development Information Service (CORDIS), 'Case studies from ESPIRIT—Information technology for business': http://cordis.europa.eu/news/rcn/7343_en.html [accessed 8 March 2018]

focus on AI (e.g. autonomous land vehicles), and to prioritise instead hardware and software with military and civil applications in the shorter term.⁶⁴⁴ This allowed the SCI to make significant contributions to supercomputing, crucial in weapons design, code breaking, and other areas of national defence.

The goals of the Japanese programme, on the contrary, remained unchanged throughout its implementation. In the 1990s, this lack of flexibility was seen as a weakness by American observers due to the changing nature of the computer industry—which meant that by the time the programme ended, some of its findings were already obsolete or no longer in demand.⁶⁴⁵ Although the programme did not produce an intelligent machine, it did manage to develop prototype computers that could function at high speeds and the software to control and program them. Overall, the Japanese programme was not considered a success by the international IT community, and indeed the US continued to be the leader in computer design and software in the 1990s.

Therefore, despite having generated a great deal of anxiety, the Japanese programme actually seems to have been the least successful of all the R&D initiatives of the 1980s. Moreover, focusing on the links between research and its application in both the civilian and military sectors appeared to yield good results. Pre-competitive research, although capable of strengthening the IT community, was not sufficient to enhance the competitiveness and performance of the UK's IT industry.

Conclusion

The idea of 'AI winters' in the UK obscures important conclusions that can be drawn from the experience with the Alvey programme. Funding for particular R&D projects does seem to have spiked and dropped at particular points, the latter often occurring after disillusionment had set in. Yet the Alvey programme's results and implementation problems show that lessons from the 1970s were not properly considered, and that projects were set up without a clear understanding of how commercial and 'public good' objectives would be achieved and sustained. This was the case especially in terms of skills, and developing sustainable SMEs. Furthermore, the UK also neglected existing assets, including a highly skilled female workforce.

Professor Lighthill's assessments, for all of the criticism which has been levelled at them since, were arguably prescient. A general-purpose system was not created in the twentieth century and it is still today not considered a feasible goal by many members of the field. In addition, many of the objectives of AI since the 1960s, such as translation and speech recognition, are only nowadays yielding concrete results, a situation that corresponds more or less with Professor Lighthill's expectations.

Moreover, Professor Lighthill's suggestion of promoting the involvement of AI research with the field of application appears to have produced satisfactory results in other countries, and it seems likely that the UK could have benefitted from a similar approach in the 1980s. Indeed, the evaluators of the Alvey programme concluded that pre-competitive research, although capable of expanding and consolidating the IT research community, had not been sufficient to improve the competitiveness of the UK's IT industry.

644 Paul Ceruzzi, 'Strategic computing: DARPA and the Quest for Machine Intelligence, 1983–1993 (review)' in *The Journal of Military History*, vol. 67, no. 3 (July 2003), pp 994–996:

645 Andrew Pollack, 'Fifth Generation' Became Japan's Lost Generation', *The New York Times* (5 June 1992): <http://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html> [accessed 8 March 2018]

Some important points about AI R&D policy emerge. First, the lack of long-term evaluation of Government-backed projects hinders policy planning. Indeed, although an evaluation was carried out shortly after the Alvey programme, there was no follow-up, and it therefore remains unclear what the long-term achievements of the Alvey programme were. Second, to avoid misunderstandings and disappointment, short and long-term objectives should be clear and explicit from the outset between the various participants and funding bodies of R&D projects. And finally, the advice of expert investigators, external to the field (such as Professor Lighthill), appears to have produced useful assessments which could have informed and improved AI policy, had it been heeded. This would suggest that in future, similar evaluations should consider comparable policy developments elsewhere in the world, and past policy successes and failures within the UK. This would enable them to present suggestions which were both well informed and impartial, which could prove invaluable when determining the future goals of AI policy and the best way to achieve them.

APPENDIX 5: NOTE OF COMMITTEE VISIT TO DEEPMIND: WEDNESDAY 13 SEPTEMBER 2017

The Select Committee on Artificial Intelligence visited DeepMind's headquarters in King's Cross, London, on 13 September 2017. The Committee met with Demis Hassabis (CEO and co-founder of DeepMind), Mustafa Suleyman (Head of Applied AI and co-founder of DeepMind) and Joe Leddam (a clinical research scientist with DeepMind Health). Lunch was provided.

Eight members of the Committee were in attendance,⁶⁴⁶ as was Dr Mateja Jamnik, Specialist Adviser to the Committee.

Members were given an overview of DeepMind's history and work, and took part in a question and answer session. Topics covered included:

- **Background to DeepMind's work:** This included discussion of DeepMind's focus (making sure they are carrying out the right research, and that such work is as effective as it can be), and of DeepMind's objectives (to solve intelligence by understanding natural intelligence and applying this knowledge to machines; to use this work to make the world a better place; and to develop the world's first general purpose learning machine that can adapt to any task). Several global challenges were highlighted (access to clean water, food waste, energy consumption and climate change), as was the potential for artificial intelligence to help address them.
- **Technical work and research:** This included discussion of DeepMind's development of AlphaGo (a computer programme which can play Go) and its successes. It was made clear that artificial intelligence developments were in their infancy and no one could fully comprehend the capabilities of the systems that could be created. A demonstration was given of DeepMind's application of a Deep Q-Network (DQN) algorithm to the arcade game Breakout. The algorithm was able to learn over hundreds of games to win more quickly than a human player. 'Blackboxing' was discussed as an important engineering challenge to address.
- **Working in the UK:** This included discussion of DeepMind's commitment to the UK, its relationships with universities and its establishment of scholarships and sponsorship programmes to support machine learning in the UK (to address the shortage of skilled workers they needed). DeepMind was engaging with the public sector to inspire other AI companies to do the same. DeepMind anticipated that artificial intelligence systems would assist the work of humans, rather than replace it. It was a concern that the technology would be prematurely regulated, which could hinder development and further research. There was a discussion around whether a kite mark scheme could be a potential first step towards an industry-led regulatory framework, with artificial intelligence products marked to show they had met an agreed standard.
- **DeepMind Energy:** Data centres around the world use up to 3% of global energy. Given this was a key resource in the development of artificial intelligence, DeepMind had sought to address this. DeepMind's first application of AI in the energy sector was to try and reduce the amount of energy used for cooling a Google data centre. The AI project led to a

⁶⁴⁶ Lord Clement-Jones (Chairman), Lord Giddens, Baroness Greener, Lord Hollick, Lord Holmes of Richmond, Viscount Ridley, Baroness Rock and Lord Swinfen.

reduction of up to 40% of the amount of energy used for cooling, and a 15% improvement in the building's overall energy efficiency. They were considering further applications of this in the context of other services.

- **DeepMind Health:** DeepMind had identified that the data used by the NHS was mostly paper-based and the process of caring for someone was an incredibly complex system. DeepMind had focused on two health problems: acute kidney injury (AKI) and eye conditions which lead to blindness. They had created an application which allowed doctors to react quickly to evidence of acute kidney injury, which is currently deployed at the Royal Free Hospital. They had worked with Moorfields Eye Hospital to improve the triage and time taken to see patients at risk of serious eye conditions. It took a year to clean and label the data required to inform the system being developed, and eye experts were used to help train the algorithm. There are promising early signs and the peer reviewed research should be published at some point in 2018.
- **Ethics:** This included discussion of DeepMind's work in creating the Partnership on AI, which is intended to be an industry forum to improve understanding of AI and establish best practice for its development. DeepMind were also establishing an internal ethics unit within the next year. DeepMind had created an Independent Review Panel for their work in healthcare as a demonstration of their commitment to transparency and because of the ethical issues raised in accessing patient data.

APPENDIX 6: NOTE OF COMMITTEE VISIT TO CAMBRIDGE: THURSDAY 16 NOVEMBER 2017

On 16 November, the Select Committee on Artificial Intelligence visited Cambridge to see the work being done there on artificial intelligence. The Committee visited Microsoft Research's Cambridge office, Prowler.io and Healx, two AI-focused start-ups, and the Leverhulme Centre for the Future of Intelligence (CFI), an interdisciplinary research group based at the University of Cambridge.

Six members of the Committee were in attendance,⁶⁴⁷ as was Dr Mateja Jamnik, Specialist Adviser to the Committee.

Microsoft Research

The Committee began its tour of Cambridge with a visit to Microsoft Research's Cambridge Office, where the Committee met with David Frank, Government Affairs Manager, Professor Christopher Bishop, Technical Fellow and Laboratory Director, and Abigail Sellen, Deputy Laboratory Director and Principal Researcher. Refreshments and lunch were provided to the Committee in the course of this meeting. Professor Bishop started by noting that Microsoft Research, the arms-length, international research wing of Microsoft, was celebrating its twentieth anniversary in Cambridge, and pre-dated the arrival of many of the other large US technology companies in the area. Professor Bishop had been at Microsoft since 1997, and became laboratory director two years ago.

Professor Bishop emphasised that the laboratory's work covered more than just machine learning, and encompassed a wide variety of academic and professional disciplines, with engineers, professional designers and social scientists among their staff. They had recently opened a wet laboratory on the site, where they were experimenting with programmable biology. Microsoft Research saw themselves as sitting between business and academia, and aimed for a collaborative and sustainable relationship with the latter—in this vein, Professor Bishop noted that they were cautious not to simply 'hoover up' top computer science academics, as this could be detrimental to the long-term teaching and training of skilled researchers which they themselves depended on. When asked about whether the hype around AI should be believed, Professor Bishop said that AI was overhyped, but that there was a transformational moment in software development occurring (similar to the moment when Moore's Law for hardware was identified and observed). In developing their own systems, Microsoft did not use their customers' data: instead they used the data generated by staff to inform AI systems such as Clutter in MS Outlook.

Abigail Sellen then gave a presentation focused on ethical aspects of their work. There were three overarching principles to this:

- partnership with, not replacement of, humans;
- putting human values at the centre of their applications; and
- a strong focus on a wide-ranging set of ethical considerations, including the preservation of human autonomy, beneficence, non-maleficence, and justice.

⁶⁴⁷ Lord Clement-Jones, Baroness Greender, Lord Hollick, Viscount Ridley, Baroness Rock and Lord St John of Bletso.

She told us that the key areas for progress at the present moment included a focus on addressing the potential power imbalances created by AI, unlocking the blackbox aspects of many AI systems by developing intelligible systems, mitigating the bias in algorithms, and democratising AI by ensuring that AI tools reach as many people as possible.

In her view, one of the central issues with computer science education was that it tended to be taught only, or primarily by, computer scientists, and did not integrate contributions from other disciplines (she noted that she herself had come from a background in cognitive psychology). This was leading to, for example, a focus on the intelligence, rather than intelligibility, of AI systems, which was becoming increasingly problematic. Some in the AI research community argue that research should be focused away from more complex, impenetrable deep learning models, towards more intelligible, additive models, and Microsoft Research was exploring this area.

Professor Iain Buchan

Professor Iain Buchan, Director of Healthcare Research, presented to the Committee on the democratisation of AI for healthcare, a shift in focus for Microsoft in recent years. Among the conventional approaches to healthcare and public health, he highlighted the fact that traditional computer modelling often became outdated very quickly, and that new healthcare technologies had tended to focus only on particular areas. Microsoft analysis suggested that many GPs around the world were still reliant on crude health ‘dashboards’, which highlighted what was going wrong with patient health, but offered no solutions for how to solve these problems.

Microsoft was focusing on a much more systematic, holistic approach to healthcare data, which aimed to use AI, alongside the integration of new sources and forms of patient data, to find new solutions to a range of diseases and health issues, and empower patients to take more control of their own personal wellbeing. Professor Buchan suggested that in practice, this might mean being able to reverse Type 2 diabetes in 5–10% of cases using only improved diet and lifestyle changes, assisted, for example, through a smartphone app. Similarly, Microsoft believed that a data-driven approach to understanding allergies could produce promising results. In order for AI and data to be used to improve healthcare, Professor Buchan told the Committee that trust was essential, and that organising principles between researchers and hospitals were needed.

InnerEye demo

Members of the Committee were shown a demonstration of Microsoft’s ‘InnerEye’ technology, which was being developed to assist oncologists in the analysis of x-ray and MRI scans. At present, these scans often had to be laboriously marked up by hand. For example, before treating cancer with targeted x-rays from a LINAC machine, a scan needs to be marked up to show both the target of the treatment, and any organs which need to be protected. This is time-consuming work normally performed by highly-qualified oncologists, and therefore also very expensive. The Committee was shown how the software analysed a test scan in 20 seconds, in a process which would normally take anywhere between 90 minutes and, in the most complex cases, two days. The researchers who had developed it noted that this software had the potential to dramatically reduce the cost of analysing scans, allowing far more scans to be taken over the course of a treatment, and for more accurately targeted treatment as a result. They also emphasised that the

software was not perfect, and would generally need to be checked and amended by oncologists, reaffirming their principle that this technology would augment, rather than replace, human workers.

In follow-up questions, the researchers revealed that as the system did not rely on deep learning, comparatively small datasets of less than 200 people could be used to train the system. The data used was generally from older datasets, with fewer issues over privacy involved. While the system was capable of learning from new data (for example improving mark-up analysis by learning from the corrections made by human oncologists) Microsoft had not pursued this, as this would have privacy implications, and new frameworks would be required.

Microsoft Research and accessibility

The last presentation from Microsoft Research focused on their work using AI to help those with disabilities. Cecily Morrison (a researcher in the Human Experience and Design group) and David Frank emphasised that this work often had more widespread utility as well, as many features could also be helpful for non-disabled people. They thought their work was fundamentally about finding new ways to add information about the world, and that it helped to bridge the last obstacles towards a fully inclusive society.

The Committee was then shown two demonstrations of AI-powered products developed to help the blind. The first was Microsoft's SeeingEye app, which uses machine learning to describe things using a smartphone's camera. A wide range of things could be described, from what an item of food was (from scanning the barcode) through to identifying particular people if they had already been pre-registered in the app. The other product was Project Torino, a physical computing system which helped teach blind primary school children the basic principles of coding. It used plastic hubs connected via cords which, when linked to a computer, could be used to code musical compositions or tell a story.

Prowler.io

The Committee then visited the offices of Prowler.io, a company founded in January 2016 and met with Vishal Chatrath, co-founder and CEO, and his team. Since then they had closed their first round of seed-funding in August 2016, and acquired over 50 staff, of 24 different nationalities, with 24 PhDs. Their founders explained that they had founded the company because they observed that most AI start-ups were focused on using AI for visual recognition and classification, a problem they believed to be largely solved. They set out to develop technology which could reliably make the millions of 'micro decisions' found in complex systems in dynamic environments in which there was often high degrees of uncertainty. In particular, they were focusing on transport, financial services and the games industry.

They identified two issues with conventional machine learning approaches:

- they relied on very large quantities of data (compared with humans, who generally do not); and
- due to the nature of deep learning systems, they were usually impenetrable (which is not normally an issue for image recognition, but is more problematic in decision-making applications).

On the second point, they emphasised their interest in the transparency of AI systems and the traceability of decisions made by them, and observed that this was not only about ethical principles, but also more mundane issues, such as the ability to acquire liability insurance for their products, a crucial consideration for real-world deployment. They were keen to move beyond the machine learning systems used today, and combined three widely used approaches (probabilistic modelling, multi-agent systems and reinforcement learning) to create their own innovative methodology, which they claimed to be the first of its kind. The aim was to build an approach to AI which would be observable, interpretable, and controllable.

The Committee was shown a case study where Prowler.io had developed software to try to model the demand for taxis across the city of Porto, Portugal, which they believed could improve efficiency across the entire system by 40%. They explained that, in their view, many attempts to model the movements of ride sharing and private-hire fleets had thus far not been very good, as the data tended to update too slowly, and many were based on the problematic assumption that more data from more driving would improve the models. In their view, this was not the case, as no model developed in this way would be able to account for very infrequent occurrences. Their approach, which integrated probabilistic modelling with real-world data, could help with this.

This highlighted one of their key objectives, which was to develop systems which would require far less data to work than those currently dependent on deep learning models. While it was always good to have more data, human beings generally did not require very much data to make a decision, and Prowler.io aimed to replicate that ability. They also noted that data, much like crude oil, needed to be refined before it could be used. In general, in the industry too much emphasis was placed on the data itself, and not enough is placed on the processes whereby it is processed and actually understood. As one of their team members put it, “we need big knowledge, not big data”.

When asked why they had decided to base themselves in Cambridge, they compared it to Silicon Valley—where, in their view, the development of AI was 80% hype and 20% technological development whereas Cambridge achieved the reverse ratio. They also explained that while the presence of Cambridge University was important, and some respected the long tradition of scientific achievement within the city, more prosaically, the density of large technology companies now resident there made a bigger difference. Each of them was taking a personal risk with the company, but if the company folded, most were confident that they could still find work elsewhere in the city. They emphasised that a liberal visa regime, and a positive, open message on the value of immigration in general was crucial in attracting skilled labour to the UK. They believed that government funding was less important, as raising private sector investment had not been very challenging. They also noted that Cambridge would need to develop its transport, housing and office infrastructure if it wanted to take full advantage of the AI boom.

Healx

The Committee visited Healx, a three-year-old health-orientated AI start-up with 15 employees and £1.8 million in investments. We met with Michale Bouskila-Chubb (Head of Business Development), Dr Ian Roberts (Head of Biometrics) and Richard Smith (Head of Software Engineering). Their focus was on using AI to combat rare diseases. While very few people worldwide were afflicted with each type of rare disease, collectively there were over 7,000 diseases which fell into this classification, with more than 350 million people suffering from them

worldwide. Given that any particular disease had so few sufferers, it was usually considered uneconomical by drugs companies to develop bespoke drugs to cure them. Healx aims to address this problem by using AI to identify drugs which have already received clinical approval, and repurpose them to treat rare diseases. They were a for-profit company, and charged subscription fees to the charities and pharmaceutical companies that they worked with. In some cases, they applied for ‘protection of use’ patents on drugs which they discover may have new applications, and then sold these licenses on to pharmaceutical companies.

Working closely with patient groups and charities, they used a mixture of computational biology and machine learning to understand rare diseases and identify drugs with relevant properties. When identifying drugs, they attempted to feed in data from a wide range of sources, from medical databases to journal articles. In one of their earliest cases, studying a disease which affected around 600 children worldwide, they identified a potentially relevant treatment, and progressed through to early-stage testing.

When the discussion moved on to the challenges they faced, they highlighted four main areas. Data was a crucial area, and they noted that data sharing could often be arduous, and that gaining access to medical data could be difficult as a small company which had not yet established its credibility. While open access publishing was a valuable resource for them, they could still only access around 40% of the relevant literature, with the rest kept behind expensive paywalls by academic publishers.

Like many other companies, they also struggled to recruit people with the necessary skills in machine learning, and when they did, these salaries could be very high. In terms of funding more generally, they believed that the start-up ecosystem was good at providing funding, but they had not been able to attract any interest from Government agencies.

Their final set of challenges related to communication, and they observed that managing expectations around AI could be difficult, as the hype that now surrounded it often led people to believe that AI worked like magic. They found it particularly important to communicate the limitations of AI when dealing with hopeful patient groups who were often desperate for cures. They also faced scepticism from within the pharmacological world, where many scientists were often critical of the prospects of AI for drug discovery.

The Leverhulme Centre for the Future of Intelligence

In the final part of the visit, the Committee were hosted by the Leverhulme Centre for the Future of Intelligence (CFI) at the University of Cambridge. A number of academics from within the Centre, alongside a small number of external experts, had been brought together to discuss the implications of AI from an interdisciplinary academic perspective, and to provide an overview of the CFI’s multiple strands of work on this theme.

Proceedings were introduced by Dr Stephen Cave, Executive Director and Senior Research Fellow at the CFI, before Professor Huw Price, Academic Director of the CFI, gave an overview of the Centre’s current work. He explained that the CFI had a number of outposts, including Imperial College and Berkley University in the US, and that they were attempting to bring together a diverse community of thinkers to discuss the implications of AI. They faced challenges in terms of the very broad range of issues, the need to cover both short and long-term issues, and the need to approach these questions in a highly interdisciplinary way which

resonated with technologists and policymakers. He further explained that the CFI supports 10 sub-projects in total, had joined the Partnership on AI, and had recently supported a number of international conference on AI, including two in Japan.

Trust and transparency

The first presentation was given by Professor Zoubin Ghahramani, Dr Adrian Weller and Dr Tameem Adel, who worked on the CFI's 'Trust and Transparency' sub-project. They emphasised the need for tools to be developed to facilitate transparency and interpretability, which fell into three broad categories:

- Tools for developers (e.g. for debugging AI systems);
- Tools for users (e.g. for use in the criminal justice system, so defendants and their lawyers can understand and challenge evidence used against them); and
- Tools for investigators and auditors for use when things go wrong.

There were many aspects that needed to be considered when developing explainable AI systems. They noted that their work often overlapped with cognitive psychology, as it was often not clear what constituted a good explanation, and this could change depending on the audience. They also emphasised the need to develop trustworthy approaches, rather than simply trust, as there would be some cases where people should exercise scepticism. Equally, there needed to be AI systems which could deal with uncertainty and understand their own limits, alert users when they did not understand an issue, and seek out new information to rectify this.

AI narratives

The next set of presentations were given by Dr Sarah Dillon, Kanta Dihal and Beth Singler, of the CFI's AI Narratives sub-project. Dr Dillon began by talking about the importance of fictional stories to policymaking and public debate in the area of AI. The real issues, in her view, were not about malevolent machines, but rather about AI systems which were incompetent, or whose values were not sufficiently aligned to society. Cultural values were often unclear, and this could pose challenges when attempting to reflect these values in AI. Ultimately, without scrutiny of these issues, AI risked replicating and perpetuating dominant narratives of the past.

Dr Dillon explained how science fiction could act as a 'repository of thought experiments' about the future, and Kanta Dihal focused on Isaac Asimov's famous Three Laws of Robotics. Though first published in a story in 1942, she briefly explained how their paradigm of dominance versus subjugation between humans and intelligent machines had shaped thinking ever since, for example forming the basis of regulations used by the US Navy. She noted that there was a certain perversity to this; Asimov's robot stories were usually based on the idea that these laws were fundamentally flawed, and explored ways in which they generated unintended consequences.

Beth Singler finished this section by talking about a series of short films about AI (*Pain in the Machine*, *Good in the Machine*, and *Ghost in the Machine*), which her team filmed to provoke debate about AI and its implications. Each film was released with surveys, which were then used to generate quantitative data about public opinions on the subjects raised.

Bad actors in AI

Dr Seán Ó hÉigeartaigh, Dr Shahar Avin and Dr Martina Kunz, from the Centre for the Study of Existential Risk, a sister organisation to the CFI, gave a brief overview of their work on ‘bad actors’ in relation to artificial intelligence. They focused on the potential misuse to which AI could be put by hackers and other individuals with malicious intent. Their sub-project began when they asked the question: what is different about AI with respect to cybersecurity, and how does it break existing security paradigms?

Among the points they covered, they mentioned the risks that AI could supercharge conventional targeted cyberattacks by allowing hackers to personalise attacks on a much greater scale than before. They also noted that researchers needed to consider the multiple uses to which their research could be put, not simply the optimistic scenarios they would like to see them used for. Finally, they discussed the dangers of an international arms race or a new Cold War developing between nations regarding the development and use of AI. Although they believed that efforts should be taken to shift the international development of AI from a competitive to a collaborative footing, overall, they were not optimistic about the possibility of international restrictions.

Kinds of intelligence

The final presentation was given by Dr Marta Halina, Dr Henry Shelvin and Professor José Hernández-Orallo, whose focus was on studying the kinds of intelligence found in the natural world, in order to map out potential or desirable directions for artificial intelligence. They observed that current understanding of intelligence was extremely limited, and that there were not any good ways of measuring it in its various forms, or benchmarks by which to assess the progress of projects like DeepMind’s AlphaGo.

APPENDIX 7: NOTE OF COMMITTEE VISIT TO BBC BLUE ROOM: MONDAY 20 NOVEMBER 2017

The Select Committee on Artificial Intelligence visited the BBC Blue Room on 20 November to see the work the BBC was doing in relation to artificial intelligence.

Six members of the Committee were in attendance,⁶⁴⁸ as was Dr Mateja Jamnik, Specialist Adviser to the Committee.

Matthew Postgate, Chief Technology and Product Officer at the BBC, and his colleagues, began by providing an overview of recent developments in artificial intelligence, and some of the challenges the BBC face in this area. Matthew noted that recruiting skilled people was a perennial challenge, with only around 10,000 people in the world estimated to be capable of programming neural networks. He also discussed the challenges posed by AI developments elsewhere in the world, especially in the USA and China, and why it was important that Europe consider its own cultural and ethical values, and develop its own AI systems which reflect these values.

The Committee was shown a number of technical demonstrations of what other media companies were doing with AI and machine learning. This included Netflix's efforts to hyper-personalise the shows they present to their customers, right down to the way in which particular shows were presented. They also demonstrated an AI service, Lyrebird, which could replicate individual voices, using a small amount of data.

The discussion moved on to what the BBC was doing with AI. The Committee was told of experiments using machine learning for end-of-show credit detection, which the BBC hoped would reduce the amount of work that humans needed to put into a time-consuming and tedious activity. The BBC was also exploring the use of AI to augment the work of human camera operators, with a system, trained on years of BBC archival footage, which could begin to select appropriate shots of certain kinds of shows. It was anticipated that, with time, these kinds of 'AI directors' could increase the productivity of the BBC considerably, allowing it to cover more events, especially on a regional or local level.

Finally, there was a short discussion about related issues, including how to address bias in datasets, the BBC's Data Science Research Partnership, which connects the BBC with universities, and the BBC's policies on user data. The BBC's approach to engaging with the public on AI was also discussed, and the BBC's efforts to lead public debate, while also providing AI-powered services which embodied an ethical approach to AI, were highlighted.

⁶⁴⁸ Lord Clement-Jones (Chairman), Baroness Grender, Lord Holmes of Richmond, Lord Levene of Portsoken, Lord Puttnam and Baroness Rock.

APPENDIX 8: NOTE OF SME ROUNDTABLE EVENT AT TECHUK: THURSDAY 7 DECEMBER 2017

The Select Committee on Artificial Intelligence held a roundtable meeting, hosted by techUK, to discuss the opportunities and challenges for small and medium-sized enterprises (SMEs) who are developing or using artificial intelligence.

Five members of the Committee were in attendance,⁶⁴⁹ as was Dr Mateja Jamnik, Specialist Adviser to the Committee.

The session was attended by representatives from: Accenture (UK) Ltd, Access Partnership, Adarga Ltd, Advanced, Baker & McKenzie LLP, Blue Prism Group Plc, Bristows LLP, BSI Group, Cisco Systems Ltd, Cloudera (UK) Ltd, CMS Cameron McKenna Nabarro Olswang LLP, DeepMind, DigitalGenius, Emeiatec, Five AI, FTI Consulting, Gavin Jones Consulting Ltd, Google, IBM United Kingdom Ltd, iPLATO Healthcare Ltd, Kensai, Nominet, Oracle Corporation UK Ltd, PI Ltd, Pivigo Ltd, SAP (UK) Ltd, SVGC Ltd, Unilink Software Ltd, Vodafone Ltd, Your.MD and techUK.

Benefits and opportunities

The group began by discussing the question: “Why has your business decided to develop and/or deploy artificial intelligence? What benefits does it offer your business and your customers?” The increasing availability of data was frequently mentioned, as was the inability of many organisations to process and understand this data, which was in turn hampering efforts to boost productivity. AI was thought to present a solution to this problem. It was also noted that in many cases, AI was being introduced ‘via the back door’, with companies signing up to products and services which happened to make use of AI, rather than explicitly seeking out ways in which they could utilise it.

Barriers to developing and deploying AI

The discussion then focused on the question: “What barriers have you encountered in trying to develop and/or deploy artificial intelligence?” Many of the businesses present had views on the accessibility of funding, and it was generally agreed that securing funding to scale up companies from around 10 to 100 employees could be challenging in the UK, as UK investors were considered more risk averse than their US counterparts.

The difficulty in attracting skilled AI developers was also highlighted, as there was a general shortage in the UK and elsewhere, and those that were to be found attracted high salary premiums. Other points which were made included the perceived inflexibility of the apprenticeship levy, and the number of businesses which were still not using today’s technology, and were therefore not in a position to adopt AI in the near future.

UK environment

The group was asked: “Is the UK a good environment for start-ups focusing on developing or deploying artificial intelligence? What could help improve the environment?” They responded by highlighting the role of the catapults, particularly the Digital Catapult, which was an important way in which start-ups could access equipment and advice. It was observed that more expertise was

⁶⁴⁹ Lord Clement-Jones (Chairman), Baroness Greender, Lord Puttnam, Baroness Rock and Lord St John of Bletso.

needed within companies in terms of what AI could do, so that clients could lead in developing AI-orientated solutions to real-world problems, rather than the current situation, where start-ups sometimes attempted to solve problems which did not actually exist.

Ethical considerations

The group then discussed the question: “What ethical considerations does your business make when considering the development or use of an artificial intelligence system?” It was emphasised that the public needed confidence in AI systems, and the decisions they made, if they were going to accept them. New frameworks were needed for explainability, and explainability also needed to be designed into systems from the start, as this capacity was difficult or impossible to retrofit into existing systems. New mechanisms would be needed for tracing liability as well.

More broadly, many at the event felt that ethical guidance, and a code of ethical practice in AI, was needed in addition to more clarity on the implications of new legislation like the GDPR. When the ethics of valuing data were brought up, it was emphasised that further guidance on appropriate sharing, and reciprocal arrangements, would be of use. They noted that many public organisations seemed not to know what kinds of data they could and could not share, and that further clarity in this area would reap dividends.

Industrial Strategy, AI Sector Deal and Government intervention

When asked what they would like to see from a Government AI Sector Deal, a number of points were raised. Many attendees thought that government procurement could be a major boost to AI start-ups in the UK if government departments could be encouraged to look to UK companies first. Currently, EU regulations limited what could be done in this area, but post-Brexit, some attendees felt that there would be opportunities to reassess this.

Most attendees also believed that the government needed to invite AI start-ups and SMEs to Whitehall more frequently than they currently do. Immigration restrictions were considered an issue, as many of the most skilled AI developers in the UK today had come from abroad, and this flow needed to be maintained. Many attendees were sharply critical of Innovate UK, and believed it was not adequately serving the AI development sector. Several attendees also discussed the need for greater co-operation between industry and academia, especially from universities outside of the more developed Russell Group, and it was felt that the Government could play a more active role here.

Despite these particular issues, overall there was a view that the Industrial Strategy was broadly a step in the right direction—what was needed now was action on these policies, within a reasonably rapid timeframe, and an understanding of who, exactly, would be held accountable for their execution.

APPENDIX 9: RECOMMENDATIONS RELEVANT TO THE GOVERNMENT'S NEW AI ORGANISATIONS

We have considered the roles of a number of organisations concerned with the development and oversight of AI and AI-specific policy. We have compiled the following table to illustrate which organisation in the UK should take the lead on implementing the relevant recommendations. Our recommendation on ensuring co-ordination, and avoiding overlaps, between these organisations (paragraph 369), should also be kept in mind while considering this below.

Organisation	Priorities
The AI Council	<p>Oversee delivery of national policy framework for AI (paragraph 367)</p> <p>Lead industry in introducing a mechanism to help raise awareness of when AI is being used to make decisions which affect people (paragraph 59)</p> <p>Identify accelerators and obstacles to the use of AI by businesses (paragraph 199)</p> <p>Help establish industry standards for the intelligibility of AI systems (paragraph 106)</p>
Centre for Data Ethics and Innovations	<p>Ensure that people whose data is overseen by data trusts are properly represented (paragraph 82)</p> <p>Produce guidance on suitable approaches to the sharing of public data (paragraph 85)</p> <p>Create tools and frameworks for data sharing, control and privacy (paragraph 87)</p> <p>Produce guidance on requirements for intelligibility of AI systems (paragraph 106)</p> <p>Introduce a cross-sector AI code (paragraph 420)</p>
The Alan Turing Institute	<p>Develop short post-graduate conversion courses to help people in other disciplines transfer to working in AI (paragraph 170)</p> <p>Establish mechanisms to encourage AI PhD applications from female and BAME candidates (paragraph 174)</p> <p>Develop advice for universities on spinning out companies (paragraph 160)</p>
Government Office for AI	<p>Prepare a national policy framework for AI (paragraph 367)</p> <p>Co-ordinate the work of new AI institutions, and existing bodies and regulators (paragraph 369)</p> <p>Create a bulletin board for AI public sector challenges (paragraph 218)</p> <p>Investigate ways of expanding access to public sector datasets (paragraph 85)</p> <p>Identify gaps in possible regulation relating to AI, and ensure the use of existing regulator's knowledge when developing any new regulation (paragraph 386)</p>

APPENDIX 10: ACRONYMS AND GLOSSARY

AI	Artificial intelligence
AKI	Acute kidney injury
API	Application Programming Interface
BEIS	Department for Business, Energy and Industrial Strategy
CBI	Confederation of British Industries
CFI	Leverhulme Centre for the Future of Intelligence
CIFAR	Canadian Institute for Advanced Research
CMA	Competition and Markets Authority
DARPA	Defence Advanced Research Projects Agency
DCMS	Department for Digital, Culture, Media and Sport
DFKI	German Research Center for Artificial Intelligence
DQN	Deep Q-network
DTP	Doctoral Training Partnership Scheme
EIS	Enterprise Investment Scheme
EPSRC	Engineering and Physical Sciences Research Council
GDPR	General Data Protection Regulation
GIB	Green Investment Bank
GM	Genetically modified
HAT	Hub of All Things
HFEA	Human Fertilisation and Embryology Authority
ICAEW	Institute of Chartered Accountants in England and Wales
ICO	Information Commissioner's Office
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet protocol
IPPR	Institute for Public Policy Research
IT	Information technology
IVF	In Vitro Fertilisation
JURI	European Parliament's Committee on Legal Affairs
Mbps	Megabits per second
MCHR	The Medicines and Healthcare products Regulatory Agency
MIT	Massachusetts Institute of Technology
MoD	Ministry of Defence
NAO	National Audit Office
NATO	North Atlantic Treaty Organisation

NGO	Non-Governmental Organisation
OECD	Organisation for Economic Co-operation and Development
ORBIT	The Observatory for Responsible Research and Innovation in ICT
R&D	Research & development
RSA	Royal Society for the encouragement of Arts, Manufacturers and Commerce
SME	Small and medium-sized enterprises
STEM	Science, technology, engineering and mathematics
TUC	Trades Union Congress
UBI	Universal basic income
UN	United Nations
VCT	Venture Capital Trust scheme