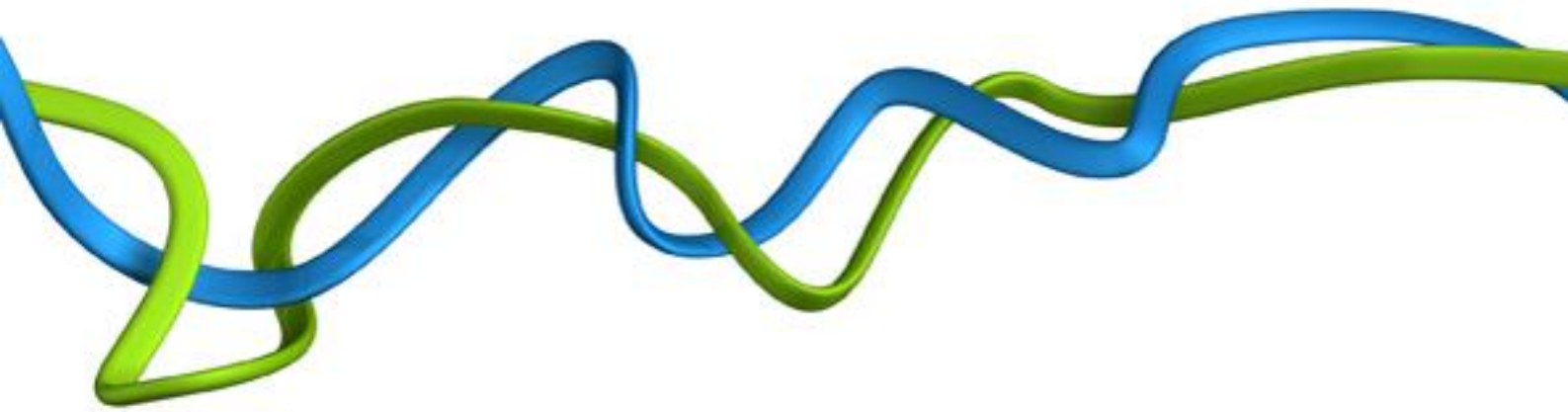
ROYAL ACADEMY OF ENGINEERING

House of Lords Select Committee on Artificial Intelligence

September 2017



Summary

1. There are many successful applications of modern artificial intelligence (AI), both in physical applications such as robotics and autonomous systems, and in non-physical applications. While current applications use 'narrow' AI that is focussed towards very specific applications and tasks, 'general' AI represents a much greater scientific and engineering challenge, although there have been some advances. In the future, the use of AI may help to tackle some of the structural and infrastructure challenges facing society. All sectors of the economy stand to benefit from artificial intelligence.

2. The acceleration of the state of the art over the past 10 years has been profound, and has a resulted in a high level of excitement that could to lead to both dystopian and utopian perspectives on what is likely to happen. Although neither of these extremes is likely to occur, continued open debate, insightful thinking and careful responses are required. Both the private sector and governments will have to move quickly to create sensible governance and control.

3. At this stage of the development of AI, it is mainly pioneering organisations that are active and experimenting. This is generating evidence of how best to create value that the fast-followers will start to exploit, at which point the use of AI will grow further, with corresponding economic benefits.

4. New types of data will continue to emerge as physical systems such as infrastructure are increasingly controlled by software and information about human activity becomes available. This will lead to new applications of AI. Organisations will need to develop the ability to make best use of their data, as well identifying opportunities to share data with other organisations.

5. As huge quantities of data are held by a few companies, there is a risk that the benefits of AI will accrue to a limited number of players if mechanisms are not put in place to encourage competition. Government needs to recognise that big data is a common good and the fuel of the AI transformation.

6. AI will impact society by displacing some jobs, enhancing human engagement in others and creating new employment and leisure time opportunities. Skills is a key issue and action needs to be taken now because of the length of the education pipeline. High-level skills are required as well as skills for people who can understand the potential of the technology in business or other areas. For the broader population, technical and data literacy should be taken as seriously as the 3Rs.

7. There are some commentators that present worst-case scenarios that capture the public's imagination and present challenges to the wider uptake of AI. A concerted public awareness campaign on both the benefits of AI and actions being taken to mitigate the downsides is needed, emphasising how AI in partnership with humans can be more productive.

8. Government, businesses and public bodies will need to consider their use of AI in decision-making, consulting widely, and ensuring that mechanisms are in place to detect and address any mistakes, biases or unintended consequences of decisions made. Ensuring transparency of algorithmic decision-making is a challenge, particularly for machine learning and self-adaptive systems.

9. While the regulatory landscape is developing, government should lead by example by applying standards to its own use of AI, to ensure accountability and help build public trust in use of algorithms. It is important to ensure that regulatory guidance and criteria are developed with sufficient expert input. The Academy stands ready to advise government on regulatory issues, as appropriate.

Introduction

10. The Royal Academy of Engineering welcomes the opportunity to respond to this call for evidence on artificial intelligence. As the UK's national academy for engineering, the Academy brings together the most successful and talented engineers from across the engineering sectors for a shared purpose: to advance and promote excellence in engineering. The Academy's response has been informed by the expertise of its Fellowship, which represents the nation's best engineering researchers, innovators, entrepreneurs, and business and industry leaders.

The pace of technological change

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

11. There are many successful applications of modern artificial intelligence (AI), both in physical applications such as robotics and autonomous systems, and in non-physical applications. Its use is growing in a wide range of sectors such as smart cities and intelligent mobility, advanced manufacturing, energy, finance, law, entertainment, education and healthcare. Additionally, companies providing AI services are emerging[1]. Novel applications are being developed where new forms of data, often in vast quantities, are being brought together with AI techniques[2]. Advanced data analytics will increasingly exploit AI technologies since the volume of data generated will require approaches that can learn patterns and highlight results autonomously – a key role for AI. Many new applications will emerge that have not yet been envisaged with the potential to provide novel solutions to many of society's major challenges. While there are pockets of innovative practice developing, the challenge will be to spread examples of early best practice.

12. The acceleration of the state of the art over the past 10 years has been profound. This is important because it means both private sector and governments will have to move quickly to create sensible governance and control. Furthermore, this governance will need to be global in nature, rather than national, which makes implementation even more difficult.

13. A number of factors have converged to make this pace of change possible, including breakthroughs in AI techniques such as neural networks, greater access to computers with significant processing power, the generation and ability to access large data sets and a greater level of investment in the technology.

14. New types of data will continue to emerge, as well as innovations in the way different datasets are used in combination. This will accelerate as associated technologies such as the Internet of Things, cloud services, personal mobile devices and data platforms develop. Correspondingly, new applications for AI will emerge that previously were not possible without this data, and without access to information about human activity or infrastructure that can be controlled by software. Goods and services that use AI in their operation are likely to be gathering their own data and combining it with data received from elsewhere. They may also share all or part of their data or results with

[1] For example, Google Deepmind are developing artificial intelligence techniques and their applications, especially in healthcare www.deepmind.com; Vivacity are applying machine learning to smart cities and intelligent mobility www.vivacitylabs.com; Feedzai are using banks' data on customer behaviour to apply predictive techniques to assess risk in real-time. Machine learning is used to detect anomalies and thus very subtle signs of fraud www.feedzai.com; CloudNC is using AI methods in manufacturing to control CNC milling machines automatically https://angel.co/cloudnc; Energi Mine uses artificial intelligence software to procure and trade energy www.energimine.com/.

[2] The Planet and Orbital Insight partnership combines Planet's satellite datasets (broad coverage, high frequency monitoring from nano satellite constellations) with Orbital's automated geo-analysis for business https://www.planet.com/pulse/planet-strikes-landmark-deal-with-orbital-insight-to-address-financial-markets/

other parties, triggering opportunities for new goods and services and related economic benefits.

15. While government has led the way with open data initiatives, much potentially valuable data remains locked away in corporate silos or within sectors[3]. Much of the issue is that this data is not curated or understood within most organisations, thus they are not able to make good use of the data they have, let alone make it available to others. The Academy is currently engaged in a study examining the opportunities and barriers for data sharing and trading based on case study research and would be happy to share early findings. Furthermore, access to large data sets is currently dominated by a small number of companies, with the risk that the benefits of AI will be held in a limited number of players.

16. Potential applications will continue to present themselves over the next 5 to 10 years, although the rate at which this happens will vary according to the specific context. For example, constraints on access to data and how it is used may be very different for healthcare applications making use of electronic patient records, say, compared to business or industry applications using corporate data – each has its own challenges. There may be varying levels of acceptability and adoption across applications which are influenced by a combination of social, economic and cultural factors. Where AI is used to control physical systems – for example, in autonomous vehicles – there are specific ethical and legal issues that could affect adoption, such as the potential shift in responsibility for safe operation from operator to designer. Cybersecurity is another challenge; breaches in security could result in a loss of integrity to algorithms as well as the data they make use of, with resulting safety and privacy implications[4].

17. The range of AI techniques is broad[5]. Current applications use 'narrow' AI that is focussed towards very specific applications and tasks. 'General' AI represents a much greater scientific and engineering challenge, although there have been some advances[6]. Active cross-disciplinary collaboration between, for example, computer scientists and neuroscientists is helping to push the state of the art, and narrow AI is beginning to apply lessons learned from one environment to another. However, one of the central challenges in achieving general AI is 'transfer learning' – the ability of computers to infer what might work in a given scenario based on knowledge gained in an apparently unrelated scenario - which is not something they currently can do. Although the timescale for general AI goes well beyond the next 20 years, it is critical to understand now how to solve the 'control problem'[7] as it is such an important issue.

18. Solving key research challenges is a high priority so that machine learning and other AI systems can be deployed safely, securely and effectively. These include ensuring outputs can be interpreted and the workings are transparent, creating systems whose behaviour can be predicted and verified with high confidence, building systems that discover causal relationships and not just correlations, and ensuring systems are not vulnerable to cyberattack[8].

19. At a national level, the use of AI may help to tackle some of the structural and infrastructure challenges resulting from, for example, climate change or population growth. It will allow us to model, understand, anticipate and respond to adverse

---

[3] Royal Academy of Engineering and IET (2015), Connecting data: driving productivity and innovation, http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity

[4] The Academy is currently engaged in an ongoing programme of work on the cyber safety and resilience of critical infrastructure and the internet of things.

[5] Artificial intelligence can be subdivided into a number of categories. Some is deterministic and its outputs predictable even if very complex. Other artificial intelligence systems are non-deterministic, learning systems and these can be further split into those such as neural networks which are frozen before release to users and those which continue to learn after release.

[6] SingularityHub (March 2017), *Google chases general intelligence with new AI that has a memory*, https://singularityhub.com/2017/03/29/google-chases-general-intelligence-with-new-ai-that-has-a-memory/

[7] Nick Bostrom (July 2014), *Superintelligence: Paths, dangers and strategies*.

[8] Royal Society (April 2017), *Machine learning: the power and promise of computers that learn by example*, https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf. Chapter 6 describes the key unsolved research challenges.

weather conditions, changing water levels, energy needs and use, terrorist plots, waste management and human safety in specific situations. All of these uses will increase our security, improve the use of resources and create stability for economic prosperity. A focus on 'benefits-led' innovation, combined with better data sharing and connectivity, will help cross-sectoral innovation to occur, resulting in the economic, social and environmental benefits that AI has the potential to bring.

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

20. Alan Turing argued[9] that there is no reason why a computer cannot perform all the functions of a human brain. Nothing has emerged since 1950 that fatally undermines Turing's argument[10]. So the current interest in AI is certainly warranted, although perhaps based on unrealistic expectations about the speed of progress.

21. Compared to previous rushes of enthusiasm for AI, this time round it is founded on practical results rather than wishful thinking. For example, the Alvey programme[11] in the 1980s was driven by the fear of a Japanese fifth generation computer threat that never materialised[12]. There is compelling evidence that current AI techniques based on access to large scale computing and information resources are applicable to a wide range of applications, in contrast to earlier approaches that delivered results that could not be transferred from one application to another. To a significant degree this is because the approach is based on automated learning rather than human-driven programming.

22. The smartphone is another technology development that is enabling the wider use of AI. This device is packed with sensors and provides a very convenient way to collect data and feed it back to cloud-based AI applications. The fact that a phone can see, hear and locate itself makes it a very powerful device. Moreover, since a large proportion of the population carriers a smartphone, AI can demand our attention extremely easily.

23. As with most technological advances, the impact in the short term will be lower than anticipated, but in the long term every device used by people, and the controls attached to infrastructure, tools, vehicles and buildings, will include some notion of adaptive and learning behaviour.

24. The level of excitement is high enough to lead to both dystopian and utopian perspectives on what is likely to happen but neither of these extremes is likely to occur. However, the existence of these perspectives reflects the profound, and very fast-moving, change that is occurring that will require continued open debate, insightful thinking and careful responses.

Impact on society

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

25. Artificial intelligence is already used routinely in everyday life, for example in intelligent personal assistants that use AI voice recognition such as Amazon's Alexa, Apple's Siri and Google's Home, and by Netflix and others for movie recommendations. It is not always visible and that will continue to be the case for a number of AI applications. If done well, most consumers will not be aware that it is AI, but only that the things they use work more effectively and consume less resource. Other applications will be more visible.

[9] in his seminal 1950 paper (Computing Machinery and Intelligence, Mind, 59 433-460

[10] Stephen Hawking and others wrote in May 2014 in The Independent that "there are no fundamental limits to what can be achieved: there is no physical law precluding particles from being organised in ways that perform even more advanced computations than the arrangements of particles in human brains".

[11] This was a British government sponsored research program in information technology that ran from 1983 to 1987, and that focused on artificial intelligence among other things.

[12] NY Times (June 1992), *'Fifth Generation' became Japan's lost generation* http://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html

26. AI will impact society by displacing some jobs, enhancing human engagement in others and creating new employment and leisure time opportunities, although there are varying views on the impact of these changes. The Industrial Digitalisation Review[13] is addressing the impact of digitalisation on industry. While there could be a displacement of jobs in the short term, industrial digitalisation has the potential to create new, better-paid jobs – both in developing suitable AI techniques and other emerging technologies, and in operating and maintaining them in particular contexts. Furthermore, the growth that would result from increased productivity and innovation could potentially lead to the creation of additional jobs. In many cases, such technologies will be used to enhance the role of humans rather than replace them. The pace of change is such that dislocations could be very painful and conventional solutions may not adequately address the problem. Some have warned against complacency by professionals such as doctors, lawyers and accountants, whose jobs could be replaced by less-expert people, new types of experts and high-performing systems[14].

27. Skills is a key issue and action needs to be taken now because of the length of the education pipeline. Many of the high-level skills required in AI – for example, for interpreting data or avoiding bias – are common with data science, and there is a shortage of such skills. There is also a skills gap for people who can work with an AI system but are not AI experts. These people understand the potential of the technology and its limitations and can see how it might be used in business, but are not in a position to advance the state of the art.

28. It will be important to include ethics in any training. The challenge of ensuring diversity and inclusion in the workforce and of reducing digital exclusion will need to be addressed. Those people whose jobs are displaced should have the opportunity to retrain. Furthermore, a much broader social and political rethink will be required that considers what 'good work' means for people, and how active and positive citizenship is recognised and rewarded in an environment where most young people currently entering the workforce will not have a long settled career.

29. While a detailed explanation of how an AI algorithm made a decision is beyond the understanding of non-experts, greater awareness of how the technology around us works is needed, and this should be addressed in schools, and in accessible ways - such as television, courses and websites - for adults. Technical and data literacy should be taken as seriously as the 3Rs since a broad population with these skills will be necessary to support industry and other economic activity.

30. The Academy is identifying the challenges of digital skills in engineering and in the workforce more broadly, and is the engineering profession's lead on diversity and inclusion, and would be happy to contribute more information in these areas.

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

31. As with other types of technical advancement, in each sector there are leading organisations, organisations that are fast-followers, those that only change when forced to, and those that refuse to change and die. At this stage of the development of AI, it is mainly pioneering organisations that are active and experimenting. This is generating evidence of how best to generate value that the fast-followers will start to exploit, at which point the use of AI will grow further, with corresponding economic benefits.

32. Successful organisations will be those that treat their data as an asset, partner with organisations in other sectors to get access to different data sources, have a positive engagement with their consumers and partners to ensure ethical concerns are

[13] Industrial Digitalisation Review - Interim Report (July 2017), Version 3.0
http://industrialdigitalisation.org.uk/wp-content/uploads/2017/07/Interim_Report_Final3_1.pdf

[14] Richard and Daniel Susskind (October 2016), *Technology will replace many doctors, lawyers, and other professionals*, Harvard Business Review, https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals

addressed and constantly monitor value versus risks. Limited access to skills could reduce the capacity of organisations to perform each of these activities. Data-centric organisations are fairly flat and laterally integrated, so that a cultural shift may be required as an organization moves away from a hierarchical command and control structure.

33. As mentioned earlier, a small number of the wealthiest companies such as Amazon, Apple, Facebook and Google also own the largest amount of data. This situation has the potential to create even greater disparities between individuals, countries and companies without mechanisms to keep them in check.

Public perception

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

34. There are some commentators that present worst-case scenarios that capture the public's imagination and present challenges to the wider uptake of AI. In the past, misinformation about genetically modified seeds impacted agriculture and lessons should be learned from this. A concerted public awareness campaign on both the benefits of AI and actions being taken to mitigate the downsides is needed, emphasising how AI in partnership with humans can be more productive.

35. There is a general lack of understanding of the different types of algorithms used in artificial intelligence and the way that they are used. The opportunities and risks associated with the use of algorithms in decision-making depend on the type of algorithm; and understanding of the context in which an algorithm functions will be essential for public acceptance and trust. Similarly, whether an AI system acts as a primary decision maker, or as an important aid and support to a human decision maker, could influence the public's understanding of and engagement with AI.

36. Any discussion with the public will need to focus on specific applications or problems that AI can solve, such as in healthcare or transport, rather than the technology in an abstract sense. It is good that the community is already setting up forums to discuss ethical issues and government and the public should engage with these. In addition, the general public needs a better understanding of concepts such as privacy verses secrecy, how to ensure cybersecurity, plus issues such as the ownership of their data and their rights associated with it.

Industry

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

37. All sectors of the economy stand to benefit from the development and use of artificial intelligence, including advanced manufacturing, built environment, energy, transport, health, aerospace and defence and insurance[15]. Many of these applications will be producing large volumes of data generated by the Internet of Things and other novel sources such as social media and crowdsourced data. Many functions that are common across multiple sectors will be impacted by AI. This will include HR, energy efficiency, logistics, business planning and customer support.

38. New business models are emerging across a number of sectors, including built environment, transport, defence and aerospace, where data underpins a service around a product or asset. For example, for smart infrastructure, pervasive monitoring and sensing strategies will generate data that enable the use of preventative maintenance strategies so that maintenance interventions are carried out when needed, rather than after a set number of hours of operation. Reliability will also be improved, as weaknesses can be detected prior to failure occurring. This could help underpin improvements in infrastructure productivity by contributing to the delivery of new

[15] Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity

infrastructure, as well as maintaining and operating existing infrastructure at highly resilient levels[16].

39. The transport sector would benefit with the adoption of 'mobility as a service', aimed at providing consumers with relevant choices in their transport solutions. Services would be provided by a multitude of transport operators, co-ordinated by one customer interface organisation that would match individual mobility needs with available transport options. An appropriate legal and regulatory environment is needed to enable this system to work, in which, the UK could take a lead if it so wishes. Machine learning techniques would be essential to the effective and efficient application of 'mobility as a service'.

40. In advanced manufacturing, AI used alongside other technologies such as big data, robotics and the Internet of Things could result in higher performance and more flexible manufacturing systems[17]. In the energy sector, such techniques could underpin greater interoperability and flexibility in an energy system that is focused on delivering services to end-users. Defence applications are also moving extremely fast, leading to challenging ethical and governance questions. Further examples are discussed in more detail in a joint report produced by the Academy and the IET in 2015[18].

41. AI technologies will also be used in autonomous systems such as autonomous vehicles, as well as those used in manufacturing, drones, maritime and space systems, and in assistive robots[19]. Both physical and non-physical applications of AI will increasingly be employed in collaboration with humans, where AI technology will act in an assistive rather than executive mode. This will necessitate robust human-centred design. A particular system design issue is how best to give the operator the right information to exercise appropriate control. In complex systems the operator may need to be highly trained to deal with decisions handed over by the AI to the human.

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

42. New types of companies have emerged, particularly in the US, whose business models are based on the aggregation of data and provision of cloud services, such as Amazon, Google (now part of Alphabet), Facebook, Microsoft and Apple. These companies hold huge quantities of data and are in a position to compete with traditional engineering sectors for a share of the market in such areas as autonomous cars and smart cities[20], as well as with other sectors such as supermarkets, finance and insurance[21].

43. The major platform vendors may monopolise data, but there are counter examples: for example, Uber has managed to collect the traffic and map data it needs to offer its services and has the potential to rival Google in the autonomous vehicle space.

44. The large data companies are themselves developing artificial intelligence capabilities in-house, in some situations by acquiring AI SME's. Google's acquisition of Deepmind is one such example. The critical mass of expertise in such an organisation is huge, and

---

[16] Royal Academy of Engineering response to the National Infrastructure Commission (15 March 2017), *National Infrastructure Commission Technology Study - call for evidence*

[17] Industrial Digitalisation Review - Interim Report (July 2017), Version 3.0 http://industrialdigitalisation.org.uk/wp-content/uploads/2017/07/Interim_Report_Final3_1.pdf

[18] Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity

[19] Royal Academy of Engineering (2015), *Innovation in autonomous systems*, http://www.raeng.org.uk/publications/reports/innovation-in-autonomous-systems

[20] Royal Academy of Engineering and IET (2015), *Connecting data: driving productivity and innovation*, http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity

[21] World Economic Forum (August 2017), *Big Tech, Not Fintech, Causing Greatest Disruption to Banking and Insurance* https://www.weforum.org/press/2017/08/big-tech-not-fintech-causing-greatest-disruption-to-banking-and-insurance

Deepmind is considerably larger than many university departments[22]. Large data companies are also competing with SMEs and other types of tech and engineering firms for graduates skilled in data science, AI, robotics and other related areas, and are able to pay large salaries[23]. The UK needs to develop a good defence against such competition, although the creation of the Alan Turing Institute has helped counter this, as has the increase in data science courses provided by UK universities[24].

45. A better form of data sharing is needed. Data is central to much of the power of AI and the large technology companies have significant advantages based on the data they hold. Finding a way to share this with others would help to level the playing field. Care must be taken, however, to preserve privacy and to comply with the General Data Protection Regulation (GDPR).

46. Trust relies on ensuring that individual, corporate and broader social benefits from data are balanced between stakeholders. There is some evidence that the public are willing to share personal data with companies to get a better service[25], but in many instances asymmetries still exist between organisations and consumers so that the organisation has a much better idea of how it can benefit from data than the consumer. There are a number of projects developing platforms[26,27,28] that allow individuals to control data securely, make it available as they see fit with safeguards and benefit directly from their personal data, responding to the need to rebalance control of data and its benefits. If data is thought of as the 'new oil', the readiness of people to give up personal data without consideration is unwise and a level playing field will not be created without addressing this.

Ethics

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

47. Individuals each have a different definition of what is ethical. Their views are formed from family, culture, education, experience, friends and colleagues. Digital technology in the past has been developed to optimise and automate standard behaviours. The increase in processing power enables these behaviours to become more customised to the situation or personal taste. Thus the variety in an individual's definition of ethical behaviour needs to be taken into account by the system designer to ensure the base premises are reasonable and there is enough flexibility to protect and respect an individual's ethical perspective[29].

48. The thinking on ethics in relation to autonomous systems is becoming increasingly well developed. For example, the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems brings together multiple voices in these communities to identify and find consensus on timely issues. It has recently published a draft document[30] on ethical concerns for AI and autonomous systems. Academy Fellows have provided evidence to this initiative and sit on an advisory group. In addition, the

---

[22] In March 2016, Deepmind employed approximately 140 researchers: http://www.techworld.com/personal-tech/google-deepmind-what-is-it-how-it-works-should-you-be-scared-3615354/

[23] A TechNation survey stated that for tech firms in the data management and analytics area, 'barriers to accessing analytical talent are currently preventing companies from reaching their full potential'. TechCity (2016), *TechNation: transforming UK industries*. http://www.techcityuk.com/wp-content/uploads/2016/02/Tech-Nation-2016_FINAL-ONLINE-1.pdf?utm_content=buffer2e58f&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

[24] RAEng and IET (2015), *Connecting data: driving productivity and innovation*, http://www.raeng.org.uk/publications/reports/connecting-data-driving-productivity

[25] A recent study of travellers' attitudes to intelligent mobility by the Transport Systems Catapult found that 57% of respondents would not mind sharing their personal data in order to get a better service.

[26] The Hub-of-All-Things, http://hubofallthings.com/.

[27] Databox project is developing a privacy-aware data analytics platform to collate, curate, and mediate access to personal data www.databoxproject.uk/

[28] Digital Prosumer is developing a platform that allows individuals to take control of the monetisation of their personal data www.digitalprosumer.co.uk

[29] Ethics for Big Data and Analytics, http://www.ibmbigdatahub.com/whitepaper/ethics-big-data-and-analytics

[30] IEEE (December 2016), Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (AI/AS), http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf

British Standards Institute has published a guide to the ethical design and application of robots and robotic systems[31].

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

49. Ensuring transparency of algorithmic decision-making is a challenge, particularly for machine learning and self-adaptive systems. Even with techniques to ensure some human supervision, it is already the case that AI is achieving good results, but without humans being able to comprehend how. Issues of governance and accountability will need to be considered in the design and development of these systems. There are many human influences in algorithmic decision-making, including setting criteria choices and optimisation functions that will need to be understood and documented. Software engineering of algorithms will need to introduce mechanisms for logging and providing feedback to allow for greater accountability.

50. The quality of the output from an algorithm depends on the availability, quality and appropriateness of the data that is fed in, as well as the fitness for purpose or 'correctness' of the algorithm itself. Transparency about the data on which the algorithmic decisions are being made is critical to ensure accountability. Good quality metadata is vital for understanding the provenance, quality and timeliness of data.

51. Government, businesses and public bodies will need to consider their use of algorithms in decision-making, consulting widely, and ensuring that mechanisms are in place to detect and address any mistakes, biases or unintended consequences of decisions made. However, the Academy recognises that there are significant implications for government, businesses and public bodies in requiring increased transparency. The regulatory context, commercial constraints, cultural attitudes and the need to protect personal data affect the willingness and ability of organisations to share information to achieve transparency. One example where transparency is required is the key science journals that require the availability of both data and algorithms before publication.

52. It will be important for data protection safeguards to be built into software and services from the earliest stages of development. In particular, the requirements of the General Data Protection Regulations will have substantial impact, as noted above. There will be requirements for systems with properties that can be checked by regulators or the public without compromising data protection. Mechanisms could include the disclosure of certain key pieces of information, including aggregate results and benchmarks, when communicating algorithmic performance to the public. Further research into effective mechanisms and strong leadership is required to address the evolving intellectual property and legal constraints.

53. In applications such as autonomous vehicles, there are technical challenges around creating fully transparent systems. While it may be possible to understand the internal workings of individual software modules and to specify how these should perform, it may not be possible to map how the performance of individual modules impacts on the overall performance of the system. However, it is the performance of the autonomous vehicle at system level that is the central concern.

The role of the Government

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

54. Government should play a role in providing support for research and development, including testbeds and demonstrators as an 'intelligent client' in public-sector procurement and in developing the necessary skills. The Academy welcomes the

[31] British Standards Institute (April 2016), BS 8611:2016 *Robots and robotic devices - guide to the ethical design and application of robots and robotic systems*.

creation of the Industrial Strategy Challenge Fund that will help support the development of technologies such as artificial intelligence.

55. How AI is used by government, business and public bodies will ultimately determine the level of regulation required for applications of this technology. Future regulations will need to be flexible enough to accommodate different requirements, data types and possible new uses of algorithms in the future, yet ensure protection and consistency in approaches.

56. While the regulatory landscape is developing, government should lead by example by applying standards to its own use of AI, to ensure accountability and help build public trust in use of algorithms. It will be important to consider the protection of personal data, auditability, and liability for harm caused by the use of algorithms.

57. It is important to ensure that regulatory guidance and criteria are developed with sufficient expert input. The Academy stands ready to advise government on regulatory issues, as appropriate. The Academy will be producing a 'challenge paper' on the regulation of autonomous systems. This study aims to support UK industry through better understanding of the key challenges facing UK-based SMEs working on the development and deployment of autonomous systems. It will identify the issues that will drive future regulations for autonomous systems, to inform industrial strategy and the Robotics and Autonomous Systems Sector Deal (the RAS Sector Deal is being led by Fellows on the working group).

58. While the extent of the future use of algorithms in decision making will differ by sector, the Academy believes that an underlying risk is the assumption that algorithms are near-perfect, or will replace humans entirely in all decision making processes. While this might be the case in some sectors, there is a risk that new applications of AI are not being introduced properly or are introduced at the behest of people who do not fully understand how AI works, its limitations, or its potential impact on society. The Academy believes that it is important for government to have an authoritative voice on these matters.

59. The Academy advises careful monitoring by government, alongside businesses and public bodies, where the use of algorithms has a greater scope to introduce or amplify biases or discrimination. This has been noted as a particular concern in financial, recruitment, legal, criminal and education sectors where algorithms may focus on specific metrics, such as age, gender or ethnicity. While this is a significant concern, it also creates the opportunity to remove existing biases by designing systems that are independent of these variables. The issue of biases is emerging as being particularly problematic, although the issue is not with the algorithms *per se*, but rather the nature and labelling of the training dataset.

60. Government needs to recognise that big data is a common good and the fuel of the AI transformation. By ensuring good structures for citizens to control their own data, while making it available for AI applications such as large-scale health applications, government will not only accelerate positive results from AI, but also encourage positive citizenship.

61. Government should also focus on big data sharing for public funded digital projects and for licensed and regulated entities, and to promote data sharing growth as a facilitator for productivity, innovation and investment.

Learning from others

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

62. The potential to create a 'data-driven' economy is affected by free flows of data across international boundaries, as well as between organisations. The European Commission

has identified sharing of data in commercial contexts as a key concern[32], and in its working paper discusses the emerging issues around the free flow of data and improved sharing of commercial data, including machine-generated data which are either non-personal in nature or personal data that have been anonymised.

63. A useful benchmark could be provided by the Finnish government's decision to introduce legislation to incentivise the use of 'mobility as a service' in the use of data, machine learning, and the frameworks to share data across traditional 'siloed sectors'.

64. It will also be important to monitor industry consortia such as Open AI and The Partnership for AI[33].

65. The European Parliament is developing its thinking on the legal framework for robotics. A draft report[34], authored by a Luxembourg MEP, outlines rules to govern how robots interact with humans and was approved by the European Parliament Committee on Legal Affairs in January 2017. This follows a recent EU project, RoboLaw[35], which investigated the ethical and legal principles raised by robotic applications and provided European and national regulators with guidelines to deal with them.

[32] European Commission (January 2017), Commission Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final, https://ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy

[33] https://www.partnershiponai.org

[34] European Parliament Committee on Legal Affairs (May 2016), *Draft report with recommendations to the Commission on Civil Law Rules on Robotics* (2015/2013(INL))

[35] EU project RoboLaw (September 2014), *Regulating emerging robotic technologies in Europe: robotics facing law and ethics*, http://www.robolaw.eu/RoboLaw_files/documents/robolaw_d6.2_guidelinesregulatingrobotics_20140922.pdf