Government
Office for Science

Artificial intelligence: opportunities and implications for the future of decision making

Government Office for Science

Foreword

We are in currently in the foothills of a new technological revolution. Artificial intelligence has the potential to be as transformative in our lifetimes as the steam-powered economy of the 19th century.

Already it’s letting us to talk to our smartphones, recommending us music, describing photos for the visually impaired and flagging up fire risks in cities.

In the near future we could see it deployed in everything from driverless cars, to intelligent energy grids, to the eradication of infectious diseases.

In government too we are looking at the potential applications of this technology in the delivery of public services.

Our Government Data Programme is increasing the number of projects and data scientists in government, while playing a leading role in establishing the appropriate use of these powerful new tools.

As one the world’s leading digital nations, artificial intelligence presents a huge opportunity for the UK.

Get this right, and we can create a more prosperous economy with better and more fulfilling jobs. We can protect our environment by using resources more efficiently. And we can make government smarter, using the power of data to improve our public services.

As we've seen already in many areas, much routine cognitive work - the filing, sifting and sorting - can increasingly be automated, freeing people up to focus on the more human aspects of any job.

The Prime Minister has announced an independent review of modern employment practices, so that the support we provide businesses and workers keeps pace with changes in the labour market and the economy.

Artificial intelligence also poses new questions about ethics and governance, the responsible use of data and strong cyber defences. To realise the full potential of this revolution, again we have to be ready with answers.

I am pleased that the Royal Society and the British Academy are conducting a review that will consider how best the UK might manage the use of artificial intelligence.

This note sets out where the science is heading, describes some of the implications for society and government, and shows how we can responsibly use this technology to improve the lives and living standards of everyone in Britain.

It is a timely and important piece of work from the Government Chief Scientific Adviser.

Matt Hancock
Minister for Digital and Culture

Contents

Introduction

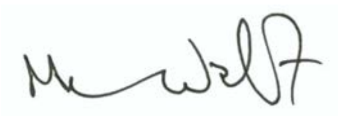
Artificial intelligence has arrived. In the online world it is already a part of everyday life, sitting invisibly behind a wide range of search engines and online commerce sites. It offers huge potential to enable more efficient and effective business and government but the use of artificial intelligence brings with it important questions about governance, accountability and ethics. Realising the full potential of artificial intelligence and avoiding possible adverse consequences requires societies to find satisfactory answers to these questions. This report sets out some possible approaches, and describes some of the ways government is already engaging with these issues.

Artificial intelligence is not a distinct technology. It depends for its power on a number of prerequisites: computing power, bandwidth, and large-scale data sets, all of which are elements of ‘big data’, the potential of which will only be realised using artificial intelligence. If data is the fuel, artificial intelligence is the engine of the digital revolution.

Much has already been written about the use of artificial intelligence and big data. This paper does not attempt to survey the whole field. Its origins lie in a seminar held at the British Academy in February 2016, chaired by Mark Walport, Government Chief Scientific Adviser and Mark Sedwill, Permanent Secretary at the Home Office, that discussed some of the legal and ethical issues around the use of artificial intelligence. The issues discussed there provide the core of this report, with additional material drawn from the views of a wide range of scientific and legal experts in the field, although we have sought to minimise detailed discussion of technical aspects in order to concentrate on the practical aspects of the debate. We hope that it serves as an introduction to the topic.

The report considers the following questions:

- What is artificial intelligence and how is it being employed?
- What benefits is it likely to bring for productivity?
- How do we best manage any ethical and legal risks arising from its use?



Sir Mark Walport
Government Chief Scientific Adviser

Mark Sedwill
Permanent Secretary, Home Office

What is artificial intelligence?

Artificial intelligence is more than the simple automation of existing processes: it involves, to greater or lesser degrees, setting an outcome and letting a computer program find its own way there. It is this creative capacity that gives artificial intelligence its power. But it also challenges some of our assumptions about the role of computers and our relationship to them.

Artificial intelligence is particularly useful for sorting data, finding patterns and making predictions. Current examples in everyday life are widespread: they include translation and speech recognition services that learn from language online, search engines that rank websites on their relevance to the user, and filters for email spam that recognise junk mail based on previous examples (see box for more uses). This list of applications is growing rapidly: artificial intelligence is enabling a new wave of innovation across every sector of the UK economy.

Artificial intelligence is a broad term (see box). More generally it refers to the analysis of data to model some aspect of the world. Inferences from these models are then used to predict and anticipate possible future events.

Some uses of artificial intelligence

Product recommendations from services such as Netflix and Amazon that evolve through users' web experiences are powered by machine learning.

The UK's 'smart motorways' use feedback on road conditions from embedded sensors and neural network systems to anticipate and manage traffic flow.

In financial markets, 'high-frequency trading' algorithms use pre-determined decision criteria to respond to market conditions many times faster than human traders are able to. Similar algorithms are being used by some financial advisors to automatically spot investment opportunities for clients.

Cornell University worked with machine learning specialists to identify the calls of right whales more accurately, making it easier to track individual whales.

Digital images from millions of satellite observations can be analysed for environmental or socio-economic trends using machine learning to identify patterns of change and development.

Statistical models are created using series of algorithms, or step-by-step instructions that computers can follow to perform a particular task. Computer algorithms are powerful tools for automating many aspects of life today, taking the step-by-step routines that underpin the administrative and operational tasks of organisations and digitising them, making them faster and more consistent. One approach to automation is to choose a series of rules to apply to inputs, leading a particular output. Most current medical self-diagnosis systems, both in books and online, use this logic. Certain combinations of answers to questions are deterministically linked to certain individual outputs. If you provide the same answers again, the algorithm will show the same result.

In recent years, however, available data and computing power have reached the point where it has become practical to develop **machine learning**: algorithms that change in response to their own output, or "computer programs that automatically improve with experience"[1].
Machine learning systems have often been shown to pick up difficult-to-spot relationships in data that may otherwise have been missed.

Most machine learning approaches are not restricted to producing a single prediction from given inputs.
Many algorithms produce probabilistic outputs, offering a range of likely predictions with associated estimates of uncertainty. The algorithms producing these probabilistic outputs are capable of being understood by humans. However, in the case of more complex machine learning systems (such as **deep learning**: see box), there are many layers of statistical operations between the input and output data. These operations have been defined by an algorithm, rather than a person. Because of this, not only is the output probabilistic, as with simpler algorithms, but the process that led to it cannot be displayed in human-understandable terms. This makes machine learning fundamentally different to the kinds of algorithms used to automate standard organisational routines.

Terminology

The range of different statistical techniques referred to here with the general term 'artificial intelligence' have emerged over a long time from many different research fields within statistics, computer science and cognitive psychology.

Consequently, authors from different disciplines tend to make different distinctions between terms like 'machine learning' and 'machine intelligence', using them to refer to related but distinct ideas.

As these techniques have been applied in different business areas they've become relevant to other tasks – so they're likely to feature also in discussions about 'data mining' and 'predictive analytics'.

While this paper looks ahead to a time when machine learning is more widespread than at present, many of the opportunities and challenges it discusses arise in other contexts too. So rather than be distracted with an academic discussion about terminology, we've chosen to use the umbrella term **artificial intelligence**.

There are many different kinds of algorithm used in machine learning. The key distinction between them is whether their learning is '**unsupervised**' or '**supervised**'.
Unsupervised learning presents a learning algorithm with an unlabelled set of data – that is, with no 'right' or 'wrong' answers – and asks it find structure in the data, perhaps by clustering elements together – for example, examining a batch of photographs of faces and learning how to say how many different people there are. Google's News service[2] uses this technique to group similar news stories together, as do researchers in genomics looking for differences in the degree to which a gene might be expressed in a given population, or marketers segmenting a target audience.

Supervised learning involves using a labelled data set to train a model, which can then be used to classify or sort a new, unseen set of data (for example, learning how to spot a particular person in a batch of photographs). This is useful for identifying elements in data (perhaps key phrases or physical attributes), predicting likely outcomes, or spotting anomalies and outliers. Essentially this approach presents the computer with a set of 'right answers' and asks it to find more of the same.

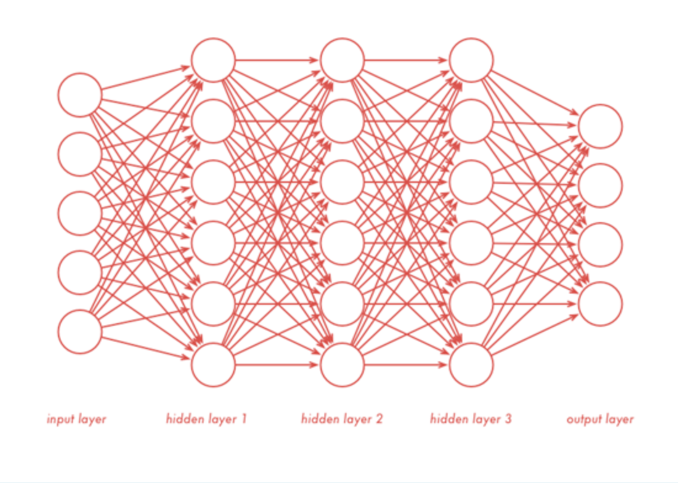
---

[1] Mitchell, T. (1997), *Machine Learning*
[2] http://news.google.com/

Current interest in machine learning is focused on '**deep learning**', a supervised learning technique combining layers of neural networks to automatically identify the features of a data set that are relevant to decision-making. Deep learning is a powerful addition to the machine learning repertoire: however, it requires very large amounts of data to be effective.
The London-based firm DeepMind (owned by Google) is a world leader in this technique.

Deep learning

Deep learning is a subset of machine learning that depends on using layers of non-linear algorithmic processes to find patterns or classify data. There are many different techniques within this general approach – but the key feature is that they each use a layered or staged design, in which outputs from the previous layer are used as inputs for the next.



Central to the interest around machine learning is the potential it offers for autonomous decision-making.
Many algorithmic processes can be used to make decisions without human input (such as when mortgage providers automatically approve lending to individuals based on their credit score).
But real autonomy comes when a system is able to learn continuously and make deductions about the world without human input. For example, self-driving cars are able to make real-time decisions about speed and direction without input from a human driver, using many interlinked machine learning systems to make sense of information about their surroundings. They are not following pre-programmed decisions but responding to changes around them.

Using autonomous decision-making in other areas of society would be a significant change in the way we use data and make choices, bringing with it important questions around accountability and trust. This is particularly true for its use by government, in light of the compact that exists between it and citizens, and its responsibility for their wellbeing and security. At the present time there is generally agreement amongst experts and policy-makers that important decisions must involve a 'human in the loop' – but the exact nature of their role, or the degree to which they influence the outcome, is something that is likely to evolve as the technology develops over time.

Artificial intelligence for innovation and productivity

Artificial intelligence holds great potential for increasing productivity, most obviously by helping firms and people use resources more efficiently, and by streamlining the way we interact with large sets of data. For example, firms like Ocado and Amazon are making use of artificial intelligence to optimise their storage and distribution networks, planning the most efficient routes for delivery and making best use of their warehousing capacity. Artificial intelligence can help firms do familiar tasks in more efficient ways. Importantly, it can also enable entirely new business models and new approaches to old problems. For example, in healthcare, data from smart phones and fitness trackers that is analysed using new machine learning techniques can improve management of chronic conditions as well as predicting and preventing acute episodes of illness.

Artificial intelligence can help both companies and individual employees to be more productive. Routine administrative and operational jobs can be learned by software agents ('bots'), which can then prioritise tasks, manage routine interactions with colleagues (or other bots), and plan schedules. Email software like Google's Smart Reply can draft messages to respondents based on previous responses to similar messages. Newsrooms are increasingly using machine learning to write sports reports and to draft articles: in the office, similar technology can produce financial reports and executive briefings.

Artificial intelligence can reduce the burden of searching large sets of data. In the legal sector, groups like ROSS, Lex Machina and CaseText are using artificial intelligence to sift court documents and legal records for case-relevant information. Other firms are using similar techniques as part of due diligence. Artificial intelligence can also offer a way of interacting with these datasets, with platforms such as IBM's Watson able to support expert systems that can answer factual natural language questions. For cybersecurity firms, artificial intelligence offers a way of recognising unusual patterns of behaviour in a network.

These examples focus on using software to do the same thing as humans but, in many cases, analysing data of volume or complexity that is beyond the analytical capability of individual humans. Indeed, artificial intelligence is not a replacement, or substitute for human intelligence. It is an entirely different way of reaching conclusions. Artificial intelligence can complement or exceed our own abilities: it can work alongside us, and even teach us, as shown by Lee Sedol's unbroken string of victories since playing AlphaGo[3]. This offers new opportunities for creativity and innovation. Perhaps the real productivity gain from artificial intelligence will be in showing us new ways to think.

The UK is a world leader in the science underpinning this technology, with a rich ecosystem of investors, employers, developers and clients, and a network of supporting bodies such as the Alan Turing Institute. Innovations developed first in universities such as Cambridge, Oxford, Imperial and University College London have found their way into tools used by millions around

---

[3] "Lee Sedol has won every single game he has played since the #AlphaGo match inc. using some new AG-like strategies - truly inspiring to see!" Demis Hassabis, CEO of DeepMind, 4th May 2016 (https://twitter.com/demishassabis/status/728020177992945664, http://gokifu.com/player/Lee+Sedol)

the globe, with increasing numbers of UK startups electing to remain in the UK, further strengthening expertise and capability in this country.

The potential is driving a rapid take-up of artificial intelligence across a range of sectors[4]. In the words of the technology pundit Kevin Kelly: “the business plans of the next 10,000 startups are easy to forecast: Take X and add AI”[5]. Estimating the size of this growth is challenging given the different definitions of ‘artificial intelligence’, ‘machine learning’ and related terms: it is also hard to specify where sectors like ‘big data’ end and ‘machine learning’ begins. But a 2015 US report[6] suggests that the global market in ‘robots and artificial intelligence-based systems will grow from $58bn in 2014 to $153bn in 2020.

Looking at big data more widely, a report earlier this year on the value of big data and the internet of things estimated £240 billion in cumulative benefits to the UK between 2015-20; manufacturing should derive the greatest benefits, with the greatest gains across all sectors set to come from efficiency savings[7]. Another report from 2015 predicts major operational savings for European governments by using big data, in addition to opportunities to increase tax revenues and reduce fraud and error[8]. A 2014 study of 500 UK businesses, meanwhile, concluded that those who make better use of customer and consumer data are between 8 and 13 per cent more productive than firms who don’t[9]. More broadly-defined forecasts for the impact of systems combining robotics, data and artificial intelligence – sometimes labelled “Industry 4.0”[10] – also promise substantial gains.

According to McKinsey, companies themselves anticipate Industry 4.0 to increase revenues by 23 per cent and productivity by 26 per cent[11]. Artificial intelligence has a central role in enabling all of this growth.

---

[4] An overview of current work is offered by Shivon Zillis of Bloomberg BETA at: www.shivonzilis.com/machineintelligence.
[5] www.wired.com/2014/10/future-of-artificial-intelligence/
[6] www.ft.com/fastft/2015/11/05/robotics-ai-become-153bn-market-20-bofa/, www.bofaml.com/content/dam/boamlimages/documents/PDFs/robotics_and_ai_condensed_primer.pdf
[7] *The Value of Big Data and the Internet of Things to the UK Economy*, CEBR for SAS, February 2016.
[8] *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011.
[9] Bakhshi, Bravo-Biosca and Mateos-Garcia (2014) – from *Data Driven Innovation* (OECD).
[10] Bundesministerium für Bildung und Forschung (2013), *Zukunftsbild “Industrie 4.0”*, PwC, *Industry 4.0: Building the digital enterprise*.
[11] McKinsey Industry 4.0 Global Expert Survey 2015.

The use of artificial intelligence by government

Government is already using data science techniques such as machine learning, and through the work of the Government Data Programme their use is growing[12]. These techniques are providing insights into a range of data, from feedback on digital service delivery to agricultural land use through the analysis satellite images. As their sophistication improves more benefits may be realised. For example, we might:

- Make existing services – such as health, social care, emergency services – more efficient by anticipating demand and tailoring services more exactly, enabling resources to be deployed to greatest effect.
- Make it easier for officials to use more data to inform decisions (through quickly accessing relevant information) and to reduce fraud and error.
- Make decisions more transparent (perhaps through capturing digital records of the process behind them, or by visualising the data that underpins a decision).
- Help departments better understand the groups they serve, in order to be sure that the right support and opportunity is offered to everyone.

Other applications will become apparent as the use of data and artificial intelligence becomes more mainstream.

Government is a special body, with unique obligations that do not fall on private organisations. It must be transparent about the way it acts, follow due process, and be accountable to its citizens. This means there are special responsibilities for government, beyond the general points outlined above, which follow from its use of artificial intelligence and big data. Recognising this, the government has published a guide to the ethical use of data science tools within government for government analysts[13]. This first iteration of a code of practice was developed with extensive external input and iterated with data scientists inside government to make it as practical and useful as possible.

Two uses with particular relevance for government are highlighted here: the use of artificial intelligence to advise, and possible legal implications of the use of artificial intelligence.

Advice

Artificial intelligence has a clear advisory role to play beyond traditional ‘decision-support systems’, supporting decision-making through assembling relevant data, identifying pertinent questions and topics for the attention of policy-makers, and in helping to generate written advice. Already, government is beginning to find value in using artificial intelligence to assist public servants in the delivery of digital services. It is likely that many types of government decisions will be deemed unsuitable to be handed over entirely to artificial intelligence systems. There will always be a ‘human in the loop’. This person's role, however, is not straightforward. If they never question the advice of the machine, the decision has *de facto* become automatic and they offer no oversight. If they question the advice they receive, however, they may be thought reckless, more so if events show their decision to be poor.

---

[12] https://data.blog.gov.uk/category/data-science/
[13] www.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication__1_.pdf

As with any adviser, the influence of these systems on decision-makers will be questioned, and departments will need to be transparent about the role played by artificial intelligence in their decisions

Legal constraints

There are currently specific legal frameworks, in addition to general legislation such as the UK Data Protection Act (1998) and the EU General Data Protection Regulation (2016), that govern the use of citizens' data by government analysts, protecting rights to privacy, ensuring equal treatment for all, and safeguarding personal identity. These are an essential ingredient in maintaining public trust in government's ability to manage data safely. Teams making use of artificial learning approaches need to understand how these existing frameworks apply in this context. For example, if deep learning is used to infer personal details that were not intentionally shared, it may not be clear whether consent has been obtained.

These current protections are effective and well-established. However, understanding the opportunities and risks associated with more advanced artificial intelligence will only be possible through trials and experimentation. For government analysts to be able to explore cutting edge techniques it may be desirable to establish sandbox areas where the potential of this technology can be investigated in a safe and controlled environment.

In addition to these three areas, the productive use of artificial intelligence in government depends on resolving wider data science issues: skills, privacy, data quality and so on. The work of the Data Science Partnership, led by the Government Digital Service (GDS), is raising the awareness of the potential of data science across government. It also provides a focal point for sharing experiences and lessons learnt to promote innovation and the spread of best practice between departments and agencies.

Effects on labour markets

The emergence of machine learning, as well as robotics, big data and autonomous systems, is likely to have significant implications for the economy and labour markets. These technologies together can be seen as part of a new wave of 'general purpose' digital technologies[14], comparable to the steam engine, and the moving assembly line, with the potential to drive significant socio-economic change. There is evidence to suggest that these technologies could drive productivity growth and so boost economic growth, but there is much uncertainty about the scale and the speed of these changes. They will depend on both the pace of technological development and the speed of its deployment by firms across the economy.

In particular, these technologies may have a particular impact on roles in the service sector, which makes up the majority of UK employment. Whilst manufacturing has been revolutionised by technological change, personal services have been less affected and, as such, have not seen the same rates of productivity growth. However, evidence from the OECD[15] indicates that the leading global service sector firms are now seeing productivity growth that outstrips their less technologically-advanced competitors.

The precise impact on labour markets of big data and robotic and autonomous systems is the subject of much debate. There is little consensus about the possible scale of job losses due to automation, which is most often the focus of these discussions. For example, one study from Deloitte found that 35% of UK jobs will be affected by automation over the next 10 to 20 years[16]. In contrast, the OECD suggest that only 10% of UK jobs are at risk[17]. It could also be that the tasks that constitute particular jobs change considerably; the same study found that underlying tasks could change considerably for a further 25% of jobs. This means that whilst a job title could stay the same, the skills needed to do it in the future could change considerably.

However, this only tells half the story and we should expect that new types of job will emerge as other jobs disappear. There are reasons to think that automation may not decrease employment, for instance, because new industries may emerge and grow as productivity gains lead to higher incomes and declining costs[18]. According to a survey by the Pew Research Centre, experts in the US are divided about the net impact of robotics and artificial intelligence: 48% think that it will displace more jobs than it creates leading to a decline in employment, 52% believe that it will create more jobs and lead to an increase in employment[19].

It is likely that automation will change the types of jobs that people do and the types of skills that they need. Evidence suggests increased automation will threaten both routine manual jobs, and routine cognitive roles[20]. Indeed, technology, coupled with trade, has already increased the proportion of high-skilled jobs and reduced the proportion of lower-middle skilled jobs[21].

---

[14] For example Brynjolfsson and McAfee, *The Second Machine Age*.
[15] www.oecd.org/employment/emp/Automation-and-independent-work-in-a-digital-economy-2016.pdf
[16] Deloitte, 2014, *Agiletown: the relentless march of technology and London's response*
[17] www.oecd.org/employment/emp/Automation-and-independent-work-in-a-digital-economy-2016.pdf
[18] www.oecd.org/employment/emp/Automation-and-independent-work-in-a-digital-economy-2016.pdf
[19] www.pewinternet.org/2014/08/06/future-of-jobs/
[20] Autor *et al.*, 2015, *Understanding U.S. Wage Inequality Trends The Review of Economics and Statistics*
[21] OECD (2015). *In It Together: Why Less Inequality Benefits All*, OECD Publishing

The increased complexity, knowledge and technological intensity of the economy is likely to lead to a continued increase in demand for high-skilled labour[22]. Automation may raise the complexity of tasks and demand higher skill levels for entry-level positions[23]. UKCES estimate that most jobs created in the decade 2012 to 2022 are expected to be high-skilled. Almost half of all employment is set to be in managerial, professional or associate professional roles by 2022[24]. Across the EU demand for skilled workers is projected to exceed supply[25]. However, some degree-level jobs involve a high proportion of routine cognitive tasks. The kinds of innovation described above mean that having a degree will not necessarily insulate an employee from the effects of automation.

Jobs that grow in the future are likely to be those that will complement technology (rather than be substituted by it). This could be because they involve skills that develop and utilise new technologies. There is significant evidence that STEM and digital skills will be increasingly in demand[26]. UKCES project that the numbers of programmers and software developers will increase by around 20% between 2012 and 2022[27].

Jobs may also complement technology if they involve tasks that are difficult to automate and there is some consensus about the types of skills these tasks will require. Frey and Osborne emphasise perception, complex manipulation, creativity and social intelligence[28]. The OECD argues that person-to-person services and occupations relying more on creativity, context adaptability, task discretion, social skills and tacit cognitive capacities have been less affected by automation[29]. The European Commission (2013) emphasise that jobs that prove resistant to automation will require people to think, communicate, organise and decide[30].

More widely, it is likely that technological change could mean that job-specific skills may perish more quickly and people may change jobs more frequently. This emphasises the need for re-skilling over the course of a career[31] and the need to be pro-active, open to change and resilient[32] . It also means that 'general purpose' skills, like problem solving and mental flexibility, that are transferrable across different domains could be increasingly valuable[33]. Government has a role to play in facilitating the development of new skills, enabling workers to retrain, either to use artificial intelligence effectively in their work or to move into areas where the value of particularly human skills—such as empathy or creativity—is evident.

---

[22] European Commission, 2013, *UK Skills Supply and Demand up to 2025*.
[23] CSIRO, 2016, *Tomorrow's Digitally Enabled Workforce*
[24] CBI, 2015, *CBI/Pearson Education and Skills Survey 2015*
[25] IPPR, 2015, *Technology, globalization and the future of work in Europe: essays on employment in a digitized economy*
[26] McKinsey Global Institute (2012), IPPR (2015), CBI (2015)
[27] UKCES, 2014, *The Future of Work: jobs and skills in 2030*.
[28] Frey and Osborne, 2013, *The Future of Employment*
[29] OECD, 2016, *The Future of Work in Figures*
[30] European Commission, 2013, *UK Skills Supply and Demand up to 2025*
[31] Deloitte, 2014, *Agiletown: the relentless march of technology and London's response*
[32] Alec Ross, 2016, *Industries of the Future*, Simon and Schuster. CSIRO, 2016, *Tomorrow's Digitally Enabled Workforce*. OECD, 2016, *The Future of Work in Figures*
[33] Autor *et al.*, 2015, *Understanding U.S. Wage Inequality Trends*, The Review of Economics and Statistics

New challenges

It is important to recognise that, alongside the huge benefits that artificial intelligence offers, there are potential ethical issues associated with some uses. Many experts feel that government has a role to play in managing and mitigating any risks that might arise. Any effort to do this would need to consider two broad areas:

- Understanding the possible impacts on individual freedoms, and on concepts such as privacy and consent, arising from the combination of machine learning approaches with the creation of ever-increasing amounts of personal data
- Adapting concepts and mechanisms of accountability for decisions made by artificial intelligence.

Statistical profiling is the use of past data to predict the likely actions or qualities of different groups. It is widely used across public and private sectors. For insurers it allows better assessment of risk. For merchants it allows better targeting of consumers. For law enforcement it allows more accurate assessment of threats.

However, profiling risks unjustly stereotyping individuals on the basis of their ethnicity, lifestyle or residence. This risk can, however, be mitigated. Organisations using these techniques in the public sector in the UK tend to avoid using race, nationality or address as criteria in order to avoid accusations of unfair discrimination. The principle of ‘innocent until proven guilty’ guides the application of these predictive techniques, which are more usually used to allocate policing resources to districts where early intervention might be beneficial, rather than being seen to target individuals. Of course, for other law enforcement agencies the ability to identify individuals accurately is precisely what is needed, helping them to avoid potentially being misled by stereotypical thinking and to make better use of resources.

Artificial intelligence techniques also make it possible to infer some kinds of private information from public data, such as the online behaviour of an individual or others linked to them (such as friends, relatives or colleagues)[34]. This information may go beyond what an individual may have originally consented to reveal. The Information Commissioner’s anonymisation code of practice[35] sets out ways for organisations to manage these risks and prevent the re-identification of individuals from combined anonymised data. As the volume of publically available data increases, however, and more powerful artificial intelligence techniques are developed, what was a ‘remote’ chance of re-identification may become more likely, and organisations will periodically need to revisit the protection they have in place.

Algorithmic bias may contribute to the risk of stereotyping. One principal source of this bias is the data used to train deep learning systems. As an example, imagine a university that uses a machine learning algorithm to assess applications for admission. The historical admissions data that is used to train the algorithm reflects the biases, conscious or unconscious, of these early admissions processes. Biases present in society can be perpetuated in this way, exacerbating unfairness. To mitigate this risk, technologists should identify biases in their data, and take steps to assess their impact.

---

[34] See, for example, Kosinski *et al*. (2013); Lindamood *et al*. (2009), Zheleva & Getoor (2009), He *et al*. (2006)
[35] https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/

These issues are the subject of current debate in university computer science departments, policy think-tanks and newspaper offices across the UK and around the world. This debate centres on the issue of governance and the sort of practical responses that might be developed by society. For example, various authors and industry groups have suggested that it could be useful to explore: certification for creators of algorithms (perhaps following the model of chartered professions; a code of conduct covering the use of artificial intelligence; clarity on liability for harm resulting from the use of artificial intelligence; an ombudsman for supporting citizen challenges to organisations using machine learning; ethical review boards capable of assessing the potential harms and benefits to society of particular applications and combinations of artificial intelligence[36], or an agreed approach to auditing processes that involve machine learning.

Some of these ideas have been developed further by industry associations or private firms. None of these are currently suggested as government policy. All of them merit further consideration and review.

Each of these approaches require addressing the question of responsibility for actions arising from the use of artificial intelligence. As we saw earlier, many artificial intelligence processes can be opaque, which makes it hard for an individual to take direct responsibility for them. But in the event that the use of artificial intelligence causes some harm there will be calls for redress and compensation. The challenge is to establish a system that can provide this.

Current approaches to liability and negligence are largely untested in this area. Asking, for example, whether an algorithm acted in the same way as a reasonable human professional would have done in the same circumstance assumes that this is an appropriate comparison to make. The algorithm may have been modelled on existing professional practice, so might meet this test by default. In some cases it may not be possible to tell that something has gone wrong, making it difficult for organisations to demonstrate they are not acting negligently or for individuals to seek redress. As the courts build experience in addressing these questions, a body of case law will develop.

Despite current uncertainty over the nature of responsibility for choices informed by artificial intelligence, there will need to be clear lines of accountability. It may be thought necessary for a chief executive or senior stakeholder to be held ultimately accountable for the decisions made by algorithms. Without a measure of this sort it is unlikely that trust in the use of artificial intelligence could be maintained. Doing this may encourage or indeed require the development of new forms of liability insurance as a necessary condition of using artificial intelligence – at least in sensitive domains.

Many experts and commentators have suggested that transparency is necessary to ensure accountability: being clear which algorithms are used, which parameters, which data, to what end, will be necessary to determine whether the technology has been used responsibly.
This presents some practical difficulties. Understanding this information and its implications requires a degree of technical skill. There are sometimes security or commercial priorities to balance against a desire for complete transparency: more prosaically, being clear about an algorithm's parameters may well see individuals and companies changing their behaviour in

[36] The House of Commons' Science and Technology Committee recommend a Council of Ethics in their report, *The Big Data Dilemma*, published 12th February 2016.

response, gaming the system. Most fundamentally, transparency may not provide the proof sought: simply sharing static code provides no assurance that it was actually used in a particular decision, or that it behaves in the wild in the way its programmers expect on a given dataset.

Computer scientists and policy specialists are exploring technical solutions to these issues of algorithmic accountability. For example, it might be possible to demonstrate ‘procedural regularity’, or the consistent application of a given algorithm[37]. Another approach might use machine learning techniques to spot inconsistent or anomalous outcomes from an algorithm’s application. Or distributed ledger technology might enable the use and effects of a particular algorithm to be tracked.

Central to the success of any technical approach to evaluating artificial intelligence will be an appreciation that looking at the algorithm alone is not as informative as examining the algorithm acting with data, which in turn will be less informative than researching the algorithm and data with the human beings the system interacts with, since it is their behaviour that generates more data and feedback (consider the impact of police officers’ behaviour on the arrest figures used as training data for predictive policing applications). Tools for analysts to assess the likely impact of algorithms they write would need to be sensitive to these real-world contexts[38]. A full appreciation of any risk will require developers to consider the wider social context of the application.

Ultimately, trying to understand specific decisions may be less productive than focussing on the outcomes and the process used to get there. Asking whether the system achieves the end it sets out to, and that this end is desirable, might be as important as understanding the technicalities of the underlying algorithm.

[37] Joshua Kroll: http://blogs.lse.ac.uk/mediapolicyproject/2016/02/10/accountable-algorithms-a-provocation/
[38] For example, the UCL STEaPP group are developing a set of tools to do this.

Public dialogue

Public trust is a vital condition for artificial intelligence to be used productively. Trust is underpinned by trustworthiness. But whilst this can be difficult to demonstrate in complex technical areas like artificial intelligence, it can be engendered from consistency of outcome, clarity of accountability, and straightforward routes of challenge.

Effective oversight will contribute to demonstrating trustworthiness. But at its core, trust is built through public dialogue. The approach taken by the Warnock review[39] leading to the establishment of the Human Fertilisation and Embryology Authority is frequently cited as an example to follow in this context: the review into in-vitro fertilisation and embryology gave a central position to the views of the public on moral and ethical questions, supporting the design of an institution able to take a sophisticated and balanced view of a complex area and so preserve public acceptance while facilitating scientific discovery.

Work is already underway to engage with the public and understand public attitudes to some of these issues, though further work will build on this. Ipsos MORI have conducted two public engagement pieces in this field – one on machine learning and another on data science more generally, conducted in partnership with government and Sciencewise[40].

There are particular aspects of artificial intelligence that may present barriers to acceptance. It may be difficult to accept, for instance, that there will be a proportion of people incorrectly profiled as a consequence of probabilistic reasoning in the same way that in medicine, diagnostic tests will always give a certain proportion of false positives and of false negatives. And the technological demands of monitoring and comprehending the actions of machine intelligence make it likely that software tools using similar technology will be necessary to ensure effective oversight. There may be concerns about the use of machines to hold machines to account, putting a technological spin on the question, “who watches the watchers?”

The public debate will need to explore, among other issues:

- how to treat different mistakes made through the use of artificial intelligence,
- how best to understand probabilistic decision-making, and
- the extent to which we should trust decisions made *without* artificial intelligence, or against the advice of artificial intelligence systems.

In the end, public trust will be maintained through demonstrating that the technology is beneficial and that safeguards work. This will require, at a minimum:

- Correctly identifying any harmful impacts of artificial intelligence.
- Formal structures and processes that enable citizen recourse to function as intended.
- Appropriate means of redress.
- Clear accountability.
- Clearly communicating the substantial benefits for society offered by artificial intelligence.

[39] www.hfea.gov.uk/docs/Warnock_Report_of_the_Committee_of_Inquiry_into_Human_Fertilisation_and_Embryology_1984.pdf
[40] www.ipsos-mori.com/researchspecialisms/socialresearch/specareas/centralgovernment/datascienceethics.aspx

Conclusion

There are already many practical efforts underway that will advance our understanding and use of artificial intelligence. Within government, GDS are leading the way in developing digital skills, establishing what responsible practice looks like through the Data Science Ethical Framework[41], and developing our institutional skill in using artificial intelligence for the benefit of the UK. And we will continue to invest in key research areas, and work with businesses to encourage inward investment, helping us to establish a global lead in the development and implementation of artificial intelligence.

Reaping the benefits of this revolution in information technology will require an approach to ethics and governance that enables innovation, builds trust among citizens, establishes a stable environment for businesses and investors, and fosters appropriate access to the data necessary for computer science to develop this technology still further. It is important that government actively works to bring this about.

The right form of governance for artificial intelligence, and indeed for the use of digital data more widely, is not self-evident. It is important to consider forms of data governance that cover all elements of the increasingly complex space, from responsibly generating data from people's behaviour to remaining accountable for autonomous software agents. Additionally, any approach adopted must be flexible, able to adapt to new uses and more advanced forms of artificial intelligence. There are many models that can be considered. But the important task is to set out what needs to be done before considering how it is to be achieved.

The Royal Society and the British Academy are investigating the new challenges of governance of data science and machine learning. Their study will consider how best the UK might respond to the broader societal and ethical issues raised. This is to be welcomed, and should offer the clearest indication of the right way forward on ethics and governance. The UK has a strong record of effective public dialogue and regulation of emerging technologies that pose difficult regulatory and ethical issues, such as embryo research, reproductive technologies and stem cell research. In this setting, the potential of artificial intelligence to contribute to our national economy and well-being will be realised.

---

[41] www.gov.uk/government/publications/data-science-ethical-framework

Annex A: Background

This note summarises the key points raised by experts on artificial intelligence and government during a number of meetings in early 2016 (see Annex B).

These meetings were held in order to:

- Better understand the nature of the legal and ethical challenges presented by the use of machine learning technologies by government
- Identify steps government can take towards addressing these challenges.

In the course of these discussions it became clear that experts felt government has to consider how it might best contribute to the safe use of artificial intelligence in wider society, as well as how best to use these technologies in its own work. Accordingly, this note summarises salient issues raised in both of these contexts.

These discussions are part of a wider conversation about how government manages and regulates the use of digital data. The intention here is to look ahead 3-5 years in order to identify potential risks and issues that can be included in the wider ongoing government conversation about data.

Annex B: Sources

The views expressed in this report are those of the Government Chief Scientific Adviser. However, alongside desk research, this paper has been informed by a number of meetings and exchanges with experts. We are grateful to them for their insights.

- A roundtable with the British Academy, attended by technologists, scientists and legal scholars
- A closed workshop organised by Nesta, attended by scientists, social researchers, professors of law, and providers of artificial intelligence to the public sector.
- An evidence-gathering session on law and governance for the Royal Society's machine learning review
- Discussions with Pia Mancini[42] (founder of DemocracyOS), Anthony Zacherzewski[43] (Director of Democratic Society), Michael Veale from the UCL STEaPP machine learning and policy project, the Royal Society machine learning team, members of the British Academy roundtable meeting, the machine intelligence group at Microsoft Research, and members of the Cambridge University Computing Laboratory.
- Comments and contributions from technical leaders within the Government Digital Service.

[42] http://democracyos.org/, www.ted.com/talks/pia_mancini_how_to_upgrade_democracy_for_the_internet_era
[43] www.demsoc.org/

This publication available from www.gov.uk/go-science

Contact us if you have any enquiries about this publication, including requests for alternative formats, at:

Government Office for Science
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000
Email: contact@go-science.gsi.gov.uk

GS/16/19