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

THE ROYAL SOCIETY

Machine learning: the power and promise of computers that learn by example
Issued: April 2017 DES4702
ISBN: 978-1-78252-259-1

This report can be viewed online at
royalsociety.org/machine-learning

Cover image

Contents

Executive summary

Machine learning is a branch of artificial intelligence that allows computer systems to learn directly from examples, data, and experience. Through enabling computers to perform specific tasks intelligently, machine learning systems can carry out complex processes by learning from data, rather than following pre-programmed rules.

Recent years have seen exciting advances in machine learning, which have raised its capabilities across a suite of applications. Increasing data availability has allowed machine learning systems to be trained on a large pool of examples, while increasing computer processing power has supported the analytical capabilities of these systems. Within the field itself there have also been algorithmic advances, which have given machine learning greater power. As a result of these advances, systems which only a few years ago performed at noticeably below-human levels can now outperform humans at some specific tasks.

Many people now interact with systems based on machine learning every day, for example in image recognition systems, such as those used on social media; voice recognition systems, used by virtual personal assistants; and recommender systems, such as those used by online retailers. As the field develops further, machine learning shows promise of supporting potentially transformative advances in a range of areas, and the social and economic opportunities which follow are significant. In healthcare, machine learning is creating systems that can help doctors give more accurate or effective diagnoses for certain conditions. In transport, it is supporting the development of autonomous vehicles, and helping to make existing transport networks more efficient. For public services it has the potential to target support more effectively to those in need, or to tailor services to users. And in science, machine learning is helping to make sense of the vast amount of data available to researchers today, offering new insights into biology, physics, medicine, the social sciences, and more.

The UK has a strong history of leadership in machine learning. From early thinkers in the field, through to recent commercial successes, the UK has supported excellence in research, which has contributed to the recent advances in machine learning that promise such potential. These strengths in research and development mean that the UK is well placed to take a leading role in the future development of machine learning. Ensuring the best possible environment for the safe and rapid deployment of machine learning will be essential for enhancing the UK's economic growth, wellbeing, and security, and for unlocking the value of 'big data'. Action in key areas – shaping the data landscape, building skills, supporting business, and advancing research – can help create this environment.

The recent success of machine learning owes no small part to the explosion of data that is available in some areas, such as image or speech recognition. This data has provided a vast number of examples, which machine learning systems can use to improve their performance. In turn, machine learning can help address the social and economic benefits expected from so-called 'big data', by extracting valuable information through advanced data analytics. Supporting the development of this function for machine learning requires an amenable data environment, based on open standards and frameworks or behaviours to ensure data availability across sectors.

As machine learning systems become more ubiquitous, or significant in certain fields, three skills needs follow. Firstly, as daily interactions with machine learning become the norm for most people, a basic understanding of the use of data and these systems will become an important tool required by people of all ages and backgrounds. Introducing key concepts in machine learning at school can help ensure this. Secondly, to ensure that a range of sectors and professions have the absorptive capacity to use machine learning in ways that are useful for them, new mechanisms are needed to create a pool of informed users or practitioners. Thirdly, further support is needed to build advanced skills in machine learning.

There is already high demand for people with advanced skills, with specialists in the field being highly sought after, and additional resources to increase this talent pool are critically needed. 'No regrets' steps in building digital literacy and informed users will also help prepare the UK for possible changes in the employment landscape, as the fields of machine learning, artificial intelligence, and robotics develop.

There is a vast range of potential benefits from further uptake of machine learning across industry sectors, and the economic effects of this technology could play a central role in helping to address the UK's productivity gap. Businesses of all sizes across sectors need to have access to appropriate support that helps them to understand the value of data and machine learning to their operations. To meet the demand for machine learning across industry sectors, the UK will need to support an active machine learning sector, which capitalises on the UK's strength in this area, and its relative international competitive advantages. The UK's start-up environment has nurtured a number of high-profile success stories in machine learning, and strategic consideration should be given to how to maximise the value of entrepreneurial activity in this space.

The Royal Society conducted research to understand the views of members of the public towards machine learning. While most people were not aware of the term, they did know of some of its applications. There was not a single common view, with attitudes, both positive and negative, varying depending on the circumstances in which machine learning was being used. Ongoing engagement with the public will be important as the field develops.

Machine learning applications can perform well at specific tasks. In many cases it can be used to augment human roles. Although it is clear that developments in machine learning will change the world of work, predicting how this will unfold is not straightforward, and existing studies differ substantially in their projections. While offering potential for new businesses or areas of the UK economy to thrive, the disruptive potential of machine learning brings with it challenges for society, and questions about its social consequences. Some of these challenges relate to the way in which new uses of data reframe traditional concepts of, for example, privacy or consent, while others relate to how people interact with machine learning systems. Careful stewardship will be needed to ensure that the productivity dividend from machine learning benefits all in society.

Machine learning is a vibrant field of research, with a range of exciting areas for further development across different methods and applications. In addition to those areas of research that address purely technical questions, there is a collection of specific research questions where progress would directly address areas of public concern around machine learning, or constraints on its wider use. Support for research in these areas can therefore help ensure continued public confidence in the deployment of machine learning systems. These areas include algorithmic interpretability, robustness, privacy, fairness, inference of causality, human-machine interaction, and security.

Recommendations

EXTRACTING VALUE FROM DATA

Creating a data environment to support machine learning

Good progress in increasing the accessibility of public sector data has positioned the UK as a leader in this area; continued efforts are needed in a new wave of 'open data for machine learning' by Government to enhance the availability and usability of public sector data, while recognising the value of strategic datasets.

In areas where there are datasets unsuitable for general release, further progress in supporting access to public sector data could be driven by creating policy frameworks or agreements which make data available to specific users under clear and binding legal constraints to safeguard its use, and set out acceptable uses. The UK Biobank demonstrates how such a framework can work. Government should further consider the form and function of such new models of data sharing.

Continuing to ensure that data generated by charity- and publicly-funded research is open by default and curated in a way that facilitates machine driven analysis will be critical in supporting wider use of research data. Where appropriate, journals should insist on this data being made available to other researchers in its original form, or via appropriate summary statistics where sensitive personal information is involved.

In designing their studies, researchers should consider future potential uses of their data, and build in the broadest consents that are ethically acceptable, and acceptable to research participants.

Research funders should ensure that data handling, including the cost of preparing data and metadata, and associated costs, such as staff, is supported as a key part of research funding, and that researchers are actively encouraged across subject areas to apply for funds to cover this. Research funders should ensure that reviewers and panels assessing grants appreciate the value of such data management.

Extending the lifecycle of open data requires open standards

New open standards are needed for data, which reflect the needs of machine-driven analytical approaches.

The Government has a key role to play in the creation of new open standards, for example for metadata. Government should explore ways of catalysing the safe and rapid delivery of these to support machine learning in the UK.

CREATING VALUE FROM MACHINE LEARNING

Human capital, and building skills at every level

Schools need to ensure that key concepts in machine learning are taught to those who will be users, developers, and citizens.

Government, mathematics and computing communities, businesses, and education professionals should help ensure that relevant insights into machine learning are built into the current education curriculum and associated enrichment activity in schools over the next five years, and that teachers are supported in delivering these activities.

In addition to the relevant areas of mathematics, computer science, and data literacy, the ethical and social implications of machine learning should be included within teaching activities in related fields, such as Personal, Social, and Health Education.

The next curriculum reform needs to consider the educational needs of young people through the lens of the implications of machine learning and associated technologies for the future of work.

An analysis of the future data science needs of students, industry, and academia should be undertaken to inform future curriculum developments.

To equip students with the skills to work with machine learning systems across professional disciplines, universities will need to ensure that course provision reflects the skills which will be needed by professionals in fields such as law, healthcare, and finance in the future. Some exposure to machine learning techniques will also be useful in many scientific activities. Professional bodies should work with universities to adjust course provision accordingly, and to ensure accreditation schemes take these future skills needs into account.

In the short term, the most effective mechanism to support a strong pipeline of practitioners in machine learning is likely to be government support for advanced courses – namely masters degrees – which those working across a range of sectors could use to pick up machine learning skills at a high level. Government should consider introducing a new funded programme of masters courses in machine learning, potentially in parallel with encouragement for approaches to training in machine learning via Massive Open Online Courses (MOOCs), with the aim of increasing the pool of informed users of machine learning.

CREATING VALUE FROM MACHINE LEARNING (CONTINUED)

Universities and funders should give urgent attention to mechanisms which will help recruit and retain outstanding research leaders in machine learning in the academic sector. This academic leadership is critical to inspiring and training the next generation of research leaders in machine learning.

In considering the allocation of additional PhD places, as announced in the Spring 2017 budget, and new fellowships across subject areas, machine learning should be considered a priority area for investment.

Because of the substantial skills shortage in this area, near-term funding should be made available so that the capacity to train UK PhD students in machine learning is able to increase with the level of demand for candidates of a sufficiently high quality. This could be supported through allocation of the expected 1000 extra PhD places, or may require additional resources.

Machine learning and the Industrial Strategy

As it considers its future approach to immigration policy, the UK must ensure that research and innovation systems continue to be able to access the skills they need. The UK's approach to immigration should support the UK's aim to be one of the best places in the world to research and innovate, and machine learning is an area of opportunity in support of this aim.

Government's proposal that robotics and AI could be an area for early attention by the Industrial Strategy Challenge Fund is welcome. Machine learning should be considered a key technology in this field, and one which holds significant promise for a range of industry sectors.

UK Research and Innovation (UKRI) should ensure machine learning is noted as a key technology in the Robotics and AI Challenge area.

In determining the shape and nature of DARPA-style challenge funding for research, Government should have regard to facilitating the spread and uptake of machine learning across sectors.

Key sectors of UK industry – as outlined in this report – have the potential to adopt machine learning and create value from its use. However, uptake across sectors is patchy, and many areas of UK industry are not yet making use of this technology. For example, in manufacturing, pharmaceuticals, the legal sector, energy, cities, and transport there are challenges suitable for intervention, and potential for machine learning to disrupt current activities. Increasing the absorptive capacity of these sectors through the Industrial Strategy Challenge Fund could help deliver the benefits of machine learning more quickly, and Government should design challenges in these areas to push forward the use of machine learning accordingly.

Government needs to provide mechanisms to support people seeking to make use of machine learning, through public support for entrepreneurism, small business, and enterprise.

Businesses need to understand the value of data analytics as a key part of business infrastructure. Government support for business should be able to provide advice and guidance about how to make best use of data, and organisations such as Growth Hubs or the Knowledge Transfer Network should ensure their business advisers are sufficiently informed about the value of data as business infrastructure to be able to provide guidance for businesses about, for example, the value of machine learning.

The Department for Business, Energy and Industrial Strategy (BEIS) should review support networks for small businesses to ensure they are able to provide advice and guidance about how to make use of machine learning, or to effectively support businesses offering machine learning products. This includes public-sector procurement processes, and the effectiveness of support for businesses using machine learning should be considered as part of the Government's review of the Small Business Research Initiative.

MACHINE LEARNING IN SOCIETY

Machine learning and the public

Continued engagement between machine learning researchers and the public is needed: those working in machine learning should be aware of public attitudes to the technology they are advancing, and large-scale programmes in this area should include funding for public engagement activities by researchers. Government could further support this through its public engagement framework programmes.

To help ensure those working in machine learning are given strong grounding in the broader societal implications of their work, postgraduate students in machine learning should pursue relevant training in ethics as part of their studies.

The implications of machine learning for governance of data use

There are governance issues surrounding the use of data, including those concerning the sources of data, and the purposes for which it is used. For this, a new framework for data governance – one that can keep pace with the challenge of data governance in the 21st century – is necessary to address the novel questions arising in the new digital environment. The form and function of such a framework is the basis of a study by the Royal Society and British Academy.

It is not appropriate to set up governance structures for machine learning per se. While there may be specific questions about the use of machine learning in specific circumstances, these should be handled in a sector-specific way, rather than via an overarching framework for all uses of machine learning; some sectors may have existing regulatory mechanisms that can manage, while in others there may not be these existing systems.

Machine learning and the future of work

Society needs to give urgent consideration to the ways in which the benefits from machine learning can be shared across society.

A NEW WAVE OF MACHINE LEARNING RESEARCH

Progress in some areas of machine learning research will impact directly on the social acceptability of machine learning in applications and hence on public confidence and trust. Funding bodies should encourage and support research applications in these areas, though not to the exclusion of other areas of machine learning research. These areas include algorithm interpretability, robustness, privacy, fairness, inference of causality, human-machine interactions, and security.

Chapter one
Machine learning

Left
Many people already interact with machine learning systems on a daily basis, for example through virtual personal assistants on smartphones. © martin-dm.

Machine learning

Machine learning is the technology that allows systems to learn directly from examples, data, and experience.

1.1 Systems that learn from data

Recent years have seen much discussion of machine intelligence, and what this means for our health, productivity, and wellbeing. In such discussion, machine learning apparently promises to save lives, address global challenges such as climate change, and add trillions of dollars to the global economy through increasing productivity; while doing so it also fundamentally changes the nature of work, and shapes, or defines, the choices people make in everyday life. Between these extremes, there lies a potentially transformative technology, which brings with it both opportunities and challenges, and whose risks and benefits need to be navigated as its use becomes more central to everyday activities.

Machine learning is the technology that allows systems to learn directly from examples, data, and experience.

If the broad field of artificial intelligence (AI) is the science of making machines smart, then machine learning is a technology that allows computers to perform specific tasks intelligently, by learning from examples. These systems can therefore carry out complex processes by learning from data, rather than following pre-programmed rules.

Recent years have seen significant advances in the capabilities of machine learning, as a result of technical developments in the field, increased availability of data, and increased computing power. As a result of these advances, systems which only a few years ago struggled to achieve accurate results can now outperform humans at specific tasks. There now exist voice and object recognition systems that can perform better than humans at certain tasks, though these benchmark tasks are constrained in nature. For example, in 2015, researchers created a machine learning system that surpassed human capabilities in a narrow range of vision-related tasks, which focused on recognising individual hand-written digits[1].

Many people now interact with machine learning-driven systems on a daily basis: in image recognition systems, such as those used to tag photos on social media; in voice recognition systems, such as those used by virtual personal assistants; and in recommender systems, such as those used by online retailers.

In addition to these current applications, the field also holds significant future potential; further applications of machine learning are already in development in a diverse range of fields, including healthcare, education, transport, and more. Machine learning could provide more accurate health diagnostics or personalised treatments, tailor classroom activities to enhance student learning, and support intelligent transport systems. It could also support scientific advances, by drawing insights from large datasets, and drive operational efficiencies across a range of industry sectors.

1. Markoff J. 2015 A learning advance in artificial intelligence rivals human abilities. *New York Times*. 10 December 2015. See https://www.nytimes.com/2015/12/11/science/an-advance-in-artificial-intelligence-rivals-human-vision-abilities.html (accessed 22 March 2017).

By increasing our ability to extract insights from ever-increasing volumes of data, machine learning could increase productivity, provide more effective public services, and create new products or services tailored to individual needs. However, in doing so it raises questions about new uses of data, and the role of intelligent computer systems in society.

Given the scale of the potential benefits from this technology, and its increasing pervasiveness, now is the time to ensure that it is developed in a way that engenders public confidence and addresses key concerns or challenges. This is not only to manage the potential risks associated with machine learning, but also to ensure that the full range of potential benefits is realised.

There is an opportunity now – where the field of machine learning is sufficiently nascent – to both shape how this technology develops, and to ensure that the UK is at the forefront of driving this development.

The UK has a strong history of research and development in AI and machine learning. In the 1950s, it was home to early thinkers in the field, with Alan Turing posing the question "can machines think?[2]" and famously establishing the Turing Test – whether a person could distinguish between answers given by a machine and a human – as a marker of machine intelligence. The UK's world-leading research centres continue to drive the development of the field. In recent years, the UK's machine learning community has also demonstrated its strength in supporting start-ups, with high-profile companies including: DeepMind, an artificial intelligence start-up acquired by Google in 2014; VocalIQ, which develops speech recognition systems and was bought by Apple in 2015; Swiftkey, a text prediction system bought by Microsoft in 2016; and Magic Pony, whose software enables processing of visual data, which was sold to Twitter in 2016[3].

The UK is therefore well placed to continue to play a leading role in the development of machine learning, and in doing so both to enjoy the economic benefits it can deliver, and to help shape the field so that it advances in ways that deliver the greatest benefits to society as a whole.

2. Turing A. 1950 Can machines think? *Mind* **59**, 433–460.

3. Digital firms making use of machine learning – amongst a suite of tools to provide new services – are also attracting similar interest. For example, in November 2016 Skyscanner, a travel search business based in Edinburgh, was acquired by Chinese travel company Ctrip in a £1.4billion deal. See, for example: BBC News. 2016 Skyscanner sold to China travel firm Ctrip in £1.4bn deal. See http://www.bbc.co.uk/news/business-38088016 (accessed 24 November 2016).

Engagement with, and contributions to, the project.

Digital interactions:
60,000[4]

Face-to-face encounters:
15,000[5]

Wider contributions:
1,500[6]

Practitioner participation:
500[7]

1.2 The Royal Society's machine learning project

Recognising the promise of this technology, in November 2015 the Royal Society launched a policy project on machine learning. This sought to investigate the potential of machine learning over the next 5 – 10 years, and the barriers to realising that potential. In doing so, the project sought to engage with key audiences – in policy communities, industry, academia, and the public – to raise awareness of machine learning, understand views held by the public and contribute to public debate about this technology, and identify the key social, ethical, scientific, and technical issues that machine learning presents.

Overseen by the project's Working Group, and in pursuit of these goals, the Royal Society convened leading thinkers and practitioners to consider the ethical, legal, scientific, and industry issues associated with machine learning. The project also supported a public dialogue exercise to investigate the public's attitudes towards this technology, using the results of this exercise to inform its policy work and future engagement.

This process of evidence gathering has identified key areas in which action is needed to help the UK reap the full benefits of machine learning:

- Enabling the use of machine learning in extracting value from data, through a data environment that draws on open standards and open data principles;
- Building a skills base and research environment that can provide the human and technical capital to both apply and further develop machine learning; and
- Creating governance systems to address the key social and ethical challenges raised by data in the 21st century.

Making progress in each of these areas now will help ensure that the benefits of machine learning are shared across society, thereby helping to avoid a potentially substantial backlash or negative reaction to this technology.

This report outlines the significance of addressing these areas in order to ensure the UK remains at the forefront of developing machine learning, sets out the actions required, and makes recommendations that can support or catalyse further activities in this field. It also notes areas in which research can both push forward the capabilities of machine learning, and address societal challenges.

4. This figure includes online viewings of Royal Society public events, and interactions with the Society's infographics (as at 31 March 2017).
5. This figure represents attendance at Royal Society public events on machine learning (as at 31 March 2017).
6. This figure includes public dialogue participants, and attendees at expert workshops held as part of the project.
7. This figure represents practitioner engagement through the project's Working Group and Review Panel members, a workshop at Neural Information Processing Systems 2016, a hackathon run in partnership with the Digital Catapult, the Royal Society's Transforming our Futures conference, the Sackler Forum, and a workshop held with the Department for Environment, Food and Rural Affairs.

1.3 What is machine learning?

Machine learning

Machine learning is a technology that allows computers to learn directly from examples and experience in the form of data. Traditional approaches to programming rely on hard-coded rules, which set out how to solve a problem, step-by-step. In contrast, machine learning systems are set a task, and given a large amount of data to use as examples of how this task can be achieved or from which to detect patterns. The system then learns how best to achieve the desired output. It can be thought of as narrow AI: machine learning supports intelligent systems, which are able to learn a particular function, given a specific set of data to learn from.

In some specific areas or tasks, machine learning is already able to achieve a higher level of performance than people. For other tasks, human performance remains much better than that of machine learning systems. For example, recent advances in image recognition have made these systems more accurate than ever before. In one image labelling challenge, the accuracy of machine learning has increased from 72% in 2010, to 96% in 2015, surpassing human accuracy at this task[8]. However, human-level performance at visual recognition in more general terms remains considerably higher than these systems can achieve.

While not approaching the human-level intelligence which is usually associated with the term AI, the ability to learn from data increases the number and complexity of functions that machine learning systems can undertake, in comparison to traditional programming methods. Machine learning can carry out tasks of such complexity that the desired outputs could not be specified in programs based on step-by-step processes created by humans. The learning element also creates systems which can be adaptive, and continue to improve the accuracy of their results after they have been deployed[9].

Machine learning lives at the intersection of computer science, statistics, and data science. It uses elements of each of these fields to process data in a way that can detect and learn from patterns, predict future activity, or make decisions.

In one image labelling challenge, the accuracy of machine learning has increased from 72% in 2010, to 96% in 2015, surpassing human accuracy at this task.

8. The Economist. 2016 From not working to neural networking. See http://www.economist.com/news/special-report/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not (accessed 22 March 2017).
9. Shalev-Shwartz S, Ben-David S. 2014 Understanding machine learning: from theory to algorithms. Cambridge, UK: Cambridge University Press.

Branches of machine learning

There are three key branches of machine learning:

- In supervised machine learning, a system is trained with data that has been labelled. The labels categorise each data point into one or more groups, such as 'apples' or 'oranges'. The system learns how this data – known as training data – is structured, and uses this to predict the categories of new – or 'test' – data.
- Unsupervised learning is learning without labels. It aims to detect the characteristics that make data points more or less similar to each other, for example by creating clusters and assigning data to these clusters.
- Reinforcement learning focuses on learning from experience, and lies between unsupervised and supervised learning. In a typical reinforcement learning setting, an agent[10] interacts with its environment, and is given a reward function that it tries to optimise, for example the system might be rewarded for winning a game. The goal of the agent is to learn the consequences of its decisions, such as which moves were important in winning a game, and to use this learning to find strategies that maximise its rewards.

When machine learning systems are deployed, there is a key distinction between offline and online learning systems:

- Offline learning systems are trained and tested in an offline setting, and the trained models are then 'frozen' before being deployed to a live setting. Any subsequent training will also be performed in an offline setting, tested, and then deployed using conventional software change management methods. This approach is more common in machine learning systems that are deployed today, because it gives an opportunity for human verification of the system, before the system interacts with any user.
- Online learning systems are also trained and tested in an offline setting before deployment, but the learning algorithms continue to be applied to the trained model after deployment. This means that the performance of the system 'in the wild' can continue to improve in real-time in response to real-world data. It also means that there is no opportunity for human checking of the consequences of updates to the model, before users are exposed to these. For example, many email spam detection systems perform online learning in response to patterns of inbound email and user feedback on the system's accuracy.

10. Russell and Norvig (2003) define an agent as "something that perceives and acts" with AI being "the study and construction of rational agents" (see footnote 24 for full reference)

Recent progress

Many of the ideas which frame today's machine learning systems are not new; the field's statistical underpinnings date back centuries, and researchers have been creating machine learning algorithms with various levels of sophistication since the 1950s. However, in recent years, there have been significant advances which have increased the accuracy and reliability of machine learning. These advances have made existing technologies, such as voice or image recognition software, more useful, and have opened the door to a wider range of potential applications.

In addition to algorithmic advances, which have increased technical capabilities, the progress made in this field owes much to the increasing availability of data and of computing power.

Almost 90% of the world's data is estimated to have been produced within the last five years[11]. This increasingly rich data environment has provided the raw material for use in training machine learning systems. If one thinks of machine learning systems as algorithms that learn from examples, there has been an explosion in some areas in the last few years in the sets of available examples on which they can be trained. In one instance of this, openly accessible material from YouTube can be used to train machine learning systems to recognise commonly occurring patterns in images, such as cats[12].

Many advanced machine learning systems require massive computing power in order to support their analytical capabilities. The increased ability of computers to process this data has also been central to supporting recent advances[13]. For example, while processors in the 1970s could carry out 92,000 instructions per second, the processors in smartphones today can carry out billions of instructions per second[14]. Following what has been called Moore's Law, the processing power of computers has vastly increased in recent decades, roughly doubling every two years[15].

The ability to process large amounts of data, and to use this to make predictions or decisions, makes machine learning a key tool in a wide range of applications, including those based on image recognition or voice recognition.

1.4 Machine learning in daily life

The term 'machine learning' is not one with high salience for the public; research by the Royal Society and Ipsos MORI showed that only 9% of people recognise it[16]. However, many people are familiar with specific applications of machine learning[17], and interact with machine learning systems every day. Common applications include commercial recommender systems, virtual personal assistants, image processing, and a range of other systems which are pervasive, without many people being aware of the intelligence under the hood.

Many people already use specific applications of machine learning every day, without being aware of the intelligence under the hood.

11. IBM. What is big data? See www.ibm.com/software/data/bigdata (accessed 6 March 2017).
12. In one study, researchers at Google created an image classification system that could learn to recognise images of cats using unlabelled data available from YouTube. See, for example: Dean J, Ng A. 2012 Using large-scale brain simulations for machine learning and A.I. *Google Official Blog*. See https://googleblog.blogspot.co.uk/2012/06/using-large-scale-brain-simulations-for.html (accessed 22 March 2017).
13. In 2000, a CPU could hold 37.5 million transistors; by 2015, a CPU could hold over 1,400 million transistors. See, for example: Moore's Law. How overall processing power for computers will double every two years. See http://www.mooreslaw.org/ (accessed 22 March 2017).
14. The Royal Society. Learning infographic: what is machine learning? See https://royalsociety.org/topics-policy/projects/machine-learning/machine-learning-infographic/ (accessed 22 March 2017).
15. Moore G. 1965 Cramming more components onto integrated circuits. *Electronics* **38**, 114.
16. Ipsos MORI. 2017 Public views of machine learning: findings from public research and engagement (conducted on behalf of the Royal Society).
17. 89% people had heard of at least one of the examples of machine learning applications used in the Ipsos MORI study carried out with the Royal Society.

While many of the high-profile advances in the field have been linked to gaming, and usually the victory of a computer over a human opponent, the applications of machine learning are much broader. Its functions include pattern recognition, anomaly detection, and clustering.

The following sections describe some of the applications of machine learning already encountered in everyday life[18]. A range of potential applications, including in healthcare, education, and transport, are discussed in Chapter 2.

Recommender systems: suggesting products or services

Recommender systems – systems that recommend products or services on the basis of previous choices – are amongst the most widely recognised application of machine learning, even if familiarity with the underlying technology is low[19].

Recommender systems use patterns of consumption, and expressed preferences, to predict which products or services are likely to be desirable to the user. It is machine learning that processes data from previous purchases, and the purchases of others, and uses this to detect patterns and make predictions.

Such systems are used in a range of online retail environments, including Amazon and Netflix. They can also be used to promote particular types of content to social media users, such as news stories that correspond to a user's areas of interest.

Organising information: search engines and spam filtering

Machine learning also helps provide the results of queries entered in internet search engines, such as Google. These systems take the words entered as part of a search, find words and phrases that have the same or highly similar meanings, and use this information to predict the right webpages to respond to that query[20].

Spam detection systems can also use machine learning to filter emails. In this application, the system is trained using a sample of documents, which are classified as spam and non-spam, to distinguish between emails and direct them to the correct folders. In this training process, the system can learn how the presence of specific words, or the names of different senders, and other characteristics, relate to whether or not the email is spam. When deployed in the live system, it uses this learning to classify new emails, refining its training when users identify incorrect classifications.

Voice recognition and response: virtual personal assistants

Natural language processing and speech recognition systems can match the patterns of sounds produced in human speech to words or phrases they have already encountered, by distinguishing between the different audio-footprints of these sounds. Having identified the words used, they can then translate this to text, or carry out commands.

18. Canonical problems in machine learning are summarised in Table 1 and Annex 1.
19. In the Royal Society's public engagement research, 66% of those surveyed had heard of 'computer programmes which show you websites or advertisements based on your web browsing habits'.
20. The Royal Society. 2015 Machine Learning Conference Report. 22 May 2015.

Until recently, voice recognition systems suffered from low levels of accuracy, which made them difficult to use in many cases. Recent advances mean that these systems can now recognise speech much more accurately, translating the data patterns encoded within sound waves to text, and carrying out the commands contained therein. As a result, many smartphones and other devices now come equipped with virtual personal assistants; applications such as Alexa, Cortana, Google Assistant, or Siri, which respond to voice commands or answer questions.

Computer vision: tagging photos and recognising handwriting

Machine learning can support advanced image recognition systems and computer vision. Such vision requires computers to be able to detect and analyse visual images, and to associate numerical or symbolic information with those images.

In social media applications, image recognition can be used to tag objects or people in photos that have been uploaded to a website. Similar image recognition systems can also be used to recognise scanned handwritten material, for example to recognise the addresses on letters or the digits on cheques.

Gaming systems, which detect movements or gestures made by users as part of their play, also use machine learning via computer vision. The system is trained to detect what a 'body' looks like, and then uses this training to interact with its users.

Machine translation: translating text into different languages

Using machine translation, computer systems are able to automatically convert text or speech from one language into another. Efforts in this field date back to at least the early 1950s[21], but, again, it is recent advances in the field that have made these techniques more broadly useful. There now exists a range of approaches to this task, including statistical, rule-based, and neural network-based techniques[22].

Today, machine translation is used in specific translation apps for mobile phones, social and traditional media, and in international organisations that need to reproduce documents in a large number of languages.

Detecting patterns: unusual financial activity

As a result of its ability to analyse large datasets, machine learning can be used to identify patterns in data which might not be picked up by human analysts.

A common application of its pattern recognition abilities is in the fraud detection systems associated with credit card use or other payment systems. Using the normal transaction data from a large number of users, algorithms are trained to recognise typical patterns of spending. Using this data for each user, it can also learn what makes a transaction more or less likely to be fraudulent, such as the location, magnitude or timing of spending activity. Then, if a user displays an unusual pattern of spending, the system can raise a flag and the activity can be queried with the user.

21. Weaver W. 1955 Machine translation of languages. Cambridge, MA: MIT Press.

22. See, for example: Le Q, Schuster M. 2016 A neural network for machine translation, at production stage. *Google Research Blog*. See https://research.googleblog.com/2016/09/a-neural-network-for-machine.html (accessed 22 March 2017).

Machine learning sits at the intersection of artificial intelligence, data science, and statistics, and has applications in robotics.

1.5 Machine learning, statistics, data science, robotics, and AI

Machine learning is closely related to the fields of statistics and data science, which provide a range of tools and methods for data analysis and inference from data. It is also related to robotics and intelligent automation. These fields help shape the context in which people relate to many machine learning applications, and inform the opportunities and challenges associated with it. Machine learning also supports progress in these fields, as an underlying technology for both AI and data science.

Data science and statistics

At its most basic level, machine learning involves computers processing a large amount of data to predict outcomes. This process of data handling and prediction has strong links to the overlapping fields of data science and statistics, which seek to extract insights from data.

Statistical approaches can inform how machine learning systems deal with probabilities or uncertainty in decision-making, while processing and analysis techniques from data science feed into machine learning. However, both of these disciplines also include areas of study which are not concerned with creating algorithms that can learn from data to make predictions or decisions. While many core concepts in machine learning have their roots in data science and statistics, some of its advanced analytical capabilities do not naturally overlap with these disciplines.

Artificial intelligence

The term 'artificial intelligence' lacks a broadly agreed definition, but has variously been described as:

- "[...automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning..." (Bellman, 1978)
- "The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990)
- "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
- "The branch of computer science that is concerned with the automation of intelligent behaviour." (Luger and Stubblefield, 1993)
- "...that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment[23]."

Informally, these definitions relate to systems that think like humans, act like humans, think rationally, or act rationally[24].

Machine learning is a method that can help achieve 'narrow AI', in the sense that many machine learning systems can learn to carry out specific functions 'intelligently'. However, these specific competencies do not match the broad suite of capabilities demonstrated by people.

23. Nilsson N. 2010 The quest for artificial intelligence: a history of ideas and achievements. Cambridge, UK: Cambridge University Press.
24. Russell S, Norvig P. 2009 Artificial intelligence: a modern approach. New Jersey, US: Prentice Hall.

In public discourse, AI is often assumed to signify intelligence with fully human capabilities. Such human-level intelligence – or artificial general intelligence – receives significant media attention, but this is still some time from being delivered, and it is not clear when this will be possible.

Robotics

The term 'robot' usually conjures the idea of something that lives in the physical world. It covers a range of different applications, whose software sophistication ranges from zero, in the case of automata, to high, when representing intelligent systems. In the context of machine learning and AI, a 'robot' typically refers to the embodied form of AI; robots are physical agents that act in the real world. These physical manifestations might have sensory inputs and abilities powered by machine learning.

The field of robotics has also made advances in recent years, as a result of improvements in sensor technologies and materials. As a result, and combined with advances in machine learning, robotic systems contribute to applications such as autonomous vehicles and drones. Potential applications can also be found in areas such as assisted living or city management. These further advances will draw from capabilities created by machine learning, such as computer vision, language processing, and human-machine interaction[25].

A further development in the field of machine learning relates to the increased use of virtual agents, or 'bots'. The term 'bot' is sometimes used to refer to an autonomous agent deployed in software. Such agents may not have a physical manifestation, but may operate autonomously in the virtual world of the internet[26].

1.6 Origins and evolution of machine learning

Despite the recent attention given to, and hype surrounding, machine learning, fundamental ideas in the field are not so new, with early papers being published over sixty years ago.

Within the last decade, even the past five years, the field of machine learning has made revolutionary advances. These advances have been driven in part by the availability of large amounts of data and the accessibility of computing power, but also underpinned by algorithmic advances achieved by revisiting and re-envisioning the simple neural networks put forward in the 1940s and 1950s. Drawing further insights from physiology and neuroscience, artificial neural networks have been created in which hundreds of layers of processing allow systems to perform more complicated tasks. These so-called deep learning techniques have been responsible for some of the more high-profile recent advances in artificial intelligence research, such as the AlphaGo system's victory over Lee Sedol, acknowledged as the strongest human player at the game of Go, in March 2016 (see Box 1).

This recent revolution means that technologies such as voice recognition or image processing, which a few years ago were performing at noticeably below-human levels, can now outperform people at some tasks.

25. Stone P *et al.* 2016 Artificial intelligence and life in 2030, one hundred year study on artificial intelligence: report of the 2015–2016 study panel. Stanford, CA: Stanford University.

26. Bots are not robots – as described above – and are noted here for convenience rather than strict classification.

BOX 1

Recent advances in machine learning: the significance of AlphaGo

Progress in AI has frequently been marked by the ability of computer systems to play – and beat humans at – different games.

In the 1950s and 1960s, Arthur Samuel, a researcher at IBM, wrote a machine learning program that could play checkers. Samuel's program determined its next move by using a search-tree to compute possible moves and evaluating the board position which resulted from each option. The machine built up an understanding of 'good' and 'bad' moves via repeated games, and used this to conduct its assessment of the state of the board. Although it never achieved expert-level play – it was characterised as better than average – Samuel's system marked a major milestone in the history of AI for its ability to learn strategies by playing against itself.

In 1997, Deep Blue became the first computer chess-playing system to beat a reigning world chess champion, with its victory over Garry Kasparov receiving significant attention. Rather than relying on a revolutionary new algorithmic approach to game-playing, Deep Blue exploited the increased computing power available in the 1990s to perform large-scale searches of potential moves – it could reportedly process over 200 million moves per second – then pick the best one.

Then in 2011, IBM's Watson was pitted against human players in the US quiz show Jeopardy, and beat two of the show's champions, Brad Rutter and Ken Jennings.

The research that underpinned these developments sought to create rule-based systems, which encoded human knowledge about how to play the game and what moves to use in different situations. Armed with this knowledge, the computer could then use advanced search or decision-tree methods to select an appropriate response for a particular configuration of pieces. Essentially, these machines made use of increasing computing power to perform complex searches in order to select their next move.

Although successful in achieving certain tasks, this approach to replicating human intelligence was limited in its scalability and transferability: a chess-playing system could not play chequers, and a system relying on these types of rules could not be scaled to more challenging or intuitive games, such as the ancient Chinese game of Go.

The game Go originated in China over 2500 years ago. It is a game with relatively simple rules – players place stones on a board, and aim to cordon off empty space to create their territory, or to capture the stones of their opponent – but it is incredibly complex, due to the huge number of potential moves. Successful Go players therefore rely on intuition or instinct to play the game, rather than a rigid set of instructions.

Creating a computer which could win at Go was seen, until recently, as an uncompleted Grand Challenge in artificial intelligence.

In 2016, Google DeepMind's AlphaGo system changed this.

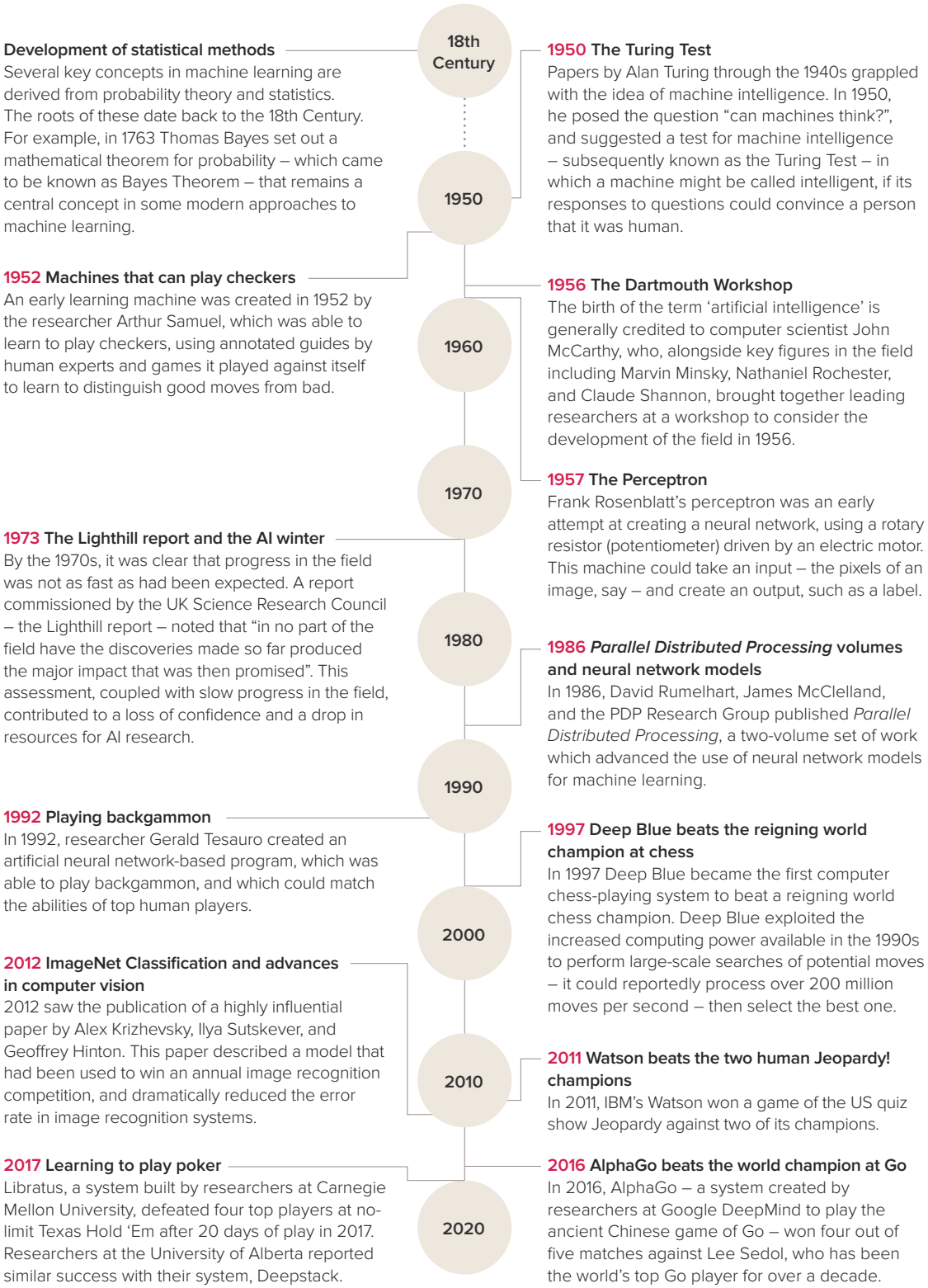
Traditional search-tree methods would not be able to process the incredibly large number of potential moves in Go. Researchers at DeepMind therefore followed a different approach: using stochastic searches and deep neural networks, they trained AlphaGo on 30 million moves from games played by humans. To further enhance its abilities, they then used reinforcement learning to allow AlphaGo to learn from thousands of games it played against itself.

This learning was put to the test in 2016, in a series of matches against Lee Sedol, who has been acknowledged as the world's top Go player for over a decade. AlphaGo played five games against Lee Sedol; it won four of them.

These victories demonstrated the ability of machine learning to tackle hugely complex tasks, and in doing so to produce solutions which humans may not have considered; pivotal moves played by AlphaGo had only a 1 in 10,000 chance of being played by a human. These were considered to be highly surprising – even beautiful – by Go experts. The AlphaGo / Lee Sedol match therefore provided a further milestone in the development of machine learning, and the history of pitching humans against machines in games to test intelligence.

FIGURE 1

Developments in machine learning and AI



1.7 Canonical problems in machine learning

Machine learning enables the analysis of data to detect patterns, and make predictions on the basis of these. The fundamental problems that it seeks to solve are summarised in Table 1.

How does machine learning work in practice? Different methods analyse data in different ways (see section 1.3), and below is one example of how machine learning can be used to detect handwriting via neural networks.

Example: neural networks for handwriting recognition

Handwriting recognition is an area in which machine learning is able to achieve high levels of accuracy.

One method of achieving this is via neural networks. Neural networks are an approach to machine learning in which layers of computational units are connected to each other in a way that is inspired by connections between neurons in the brain. One layer of these – the input units – is designed to receive information from the outside world, while the other side of the network, an output layer, communicates a decision about the data that has been received. Between these, other layers communicate information about elements of the input to each other, which contribute to the output.

In handwriting recognition, individual characters are recognised via a system known as feature extraction, which learns what letters look like by identifying the elements that make up each character. For example, if there is one vertical line at ninety degrees to a shorter horizontal line, this is very likely to be 'L'. By creating this type of rule for each character, a system is able to learn the key features that make up individual letters, and hence recognise each written character via its component features.

To enable this type of feature recognition, a neural network can be trained with a large number of examples of written text. Once it has been trained, the system can be presented with a new piece of text and, using its training, it will detect the relevant features and use these to make a decision about which letter is in front of it.

During the training phase, the accuracy of the system is improved through a process called backpropagation. This compares the output calculated by the system (the letter it predicts) to the 'true' output (defined by the user), calculates the difference between the two, and adjusts the weights between its units to improve its accuracy.

Limitations of existing approaches

While the significant progress made in recent years has enabled many impressive advances, machine learning remains subject to a number of limitations on its use. For example:

- Some approaches to machine learning rely on the accessibility of large amounts of labelled training data, the creation or curation of which can be resource-intensive, and time-consuming.
- It is difficult to develop systems with contextual understanding of a problem, or "common sense". When our expertise fails, humans fall back on common sense and will often take actions, which while not optimal, are unlikely to cause significant damage. Current machine learning systems do not define or encode this behaviour meaning that when they fail, they may fail in a serious or brittle manner.
- Humans are good at transferring ideas from one problem domain to another. This remains challenging for computers even with the latest machine learning techniques.
- Related to our failure to transfer information between problem domains is the challenge of interpretability. This can be seen as the need to represent knowledge encoded in the learning system in a form that is easily digested by humans.
- There are many constraints on the real world that we know from natural laws (such as physics) or mathematical laws such as logic. It is not straightforward to include these constraints with machine learning methods. Encoding such constraints could allow us to be more data efficient in our learning.
- Understanding the intent of humans is highly complex, it requires a sophisticated understanding of us. Current methods have a limited understanding of humans that is restricted to particular domains. This will present challenges in, for example, collaborative environments like robot helpers or even the domain of driverless cars.

In some of these areas, it is possible that technical advances will help directly address these limitations (see chapter 6).

TABLE 1

Canonical problems in machine learning

Canonical problem	Question	Some examples of applications
Classification	To which category does this data point belong?	Medical diagnosis: does this tissue show signs of disease? Banking: is this transaction fraudulent? Computer vision: what type of object is in this picture? Is it a person? Is it a building?
Regression	Given this input from a dataset, what is the likely value of a particular quantity?	Finance: what is the value of this stock going to be tomorrow? Housing: what would the price of this house be if it were sold today? Food quality: how many days before this strawberry is ripe? Image processing: how old is the person in this photo?
Clustering	Which data points are similar to each other?	E-commerce: which customers are exhibiting similar behaviour to each other, how do they group together? Video Streaming: what are the different types of video genres in our catalogue, and which videos are in the same genre?
Dimensionality reduction	What are the most significant features of this data and how can these be summarised?	E-commerce: what combinations of features allow us to summarise the behaviour of our customers? Molecular biology: how can scientists summarise the behaviour of all 20,000 human genes in a particular diseased tissue?
Semi-supervised learning	How can labelled and unlabelled data be combined?	Computer vision: how can object detection be developed, with only a small training data set? Drug discovery: which of the millions of possible drugs could be effective against a disease, given we have so far only tested a few?
Reinforcement learning	What actions will most effectively achieve a desired endpoint?	Robots: how can a robot move through its environment? Games: which moves were important in helping the computer win a particular game?

Chapter two
Emerging applications of machine learning

Left
Machine learning has a wide range of potential applications across sectors. One such application is in healthcare, where machine learning can be used to detect signs of disease from images of cells or scans, leading to more accurate diagnostic tools.

Emerging applications of machine learning

While machine learning is already supporting a range of systems in common use – as described earlier – its potential reaches much further. In areas from healthcare to education, and transport to social services, there are signs that machine learning could support improvements to the effectiveness of products or services, through increased precision or better tailoring of interventions.

In a range of industries – where there is sufficient data available to enable machine learning methods to be developed and put to use, where this data is used effectively, and where there is access to sufficient computing power – machine learning could support a step change in the delivery of products or services over the next 5 – 10 years.

As a technology with disruptive potential, machine learning could change how businesses are organised or otherwise influence the business models used in many fields. Key to this disruptive potential is the speed of change in some fields, while in other areas there will be more gradual improvements due to machine learning.

This section gives a sense of some of the applications which may be developed in the near-term, without seeking to be exhaustive[27].

2.1 Potential near-term applications in the public and private sectors

Healthcare

In healthcare, machine learning could help provide more accurate diagnoses and more effective healthcare services, through advanced analysis that improves decision-making.

One example of this function comes from breast cancer diagnosis. Breast cancer diagnoses typically include an assessment by pathologists of a tissue sample, in which doctors look for certain features that indicate the presence or extent of disease. A machine learning system trained on tissue images was able to achieve a higher accuracy than pathologists, by finding and utilising features of the image that were predictive but had not previously been used in the pathology assessments[28]. In doing so, the system was able to help doctors more accurately assess a patient's prognosis.

Another example of this function comes from the diagnosis of diabetic eye disease, which is frequently identified via specialist examination of pictures of the back of the eye[29]. The presence – or severity – of disease is determined by the presence of features in these images that indicate bleeding or fluid leakage. Researchers at Google have created a deep learning algorithm that can analyse these images[30], training the system using a

27. In this report, example applications have been selected on the basis of their recognisability in everyday life. A range of other applications exist, for example in the defence sector or financial trading. While these may raise societal or ethical challenges similar to those noted in the report, these are not the focus of this report.

28. Beck A *et al.* 2011 Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* **3**, 108. (doi: 10.1126/scitranslmed.3002564)

29. Corrado G. 2017 Applied machine learning at Google (talk at the Sackler Forum on the Frontiers of Machine Learning). See http://www.nasonline.org/programs/sackler-forum/frontiers-machine-learning.html (accessed 22 March 2017).

30. Gulshan V *et al.* 2016 Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. **312**, 2402–2410. (doi: 10.1001/jama.2016.17216)

dataset of 128,000 images, which had already been evaluated by human experts. The resulting system could diagnose the disease to a level of accuracy that was on-par with human ophthalmologists. Further work in this area is now assessing how the system could help doctors, or be evaluated in clinical studies[31]. Similar successes have been found in the use of machine learning to diagnose skin cancer[32].

There are also other machine learning techniques that can provide decision-making support for doctors. For example, IBM's 'Watson' uses machine learning in various ways. One of these is natural language processing – the form of machine learning which allows computers to process written or verbal information – which Watson uses to extract information from the vast collection of published research papers and case reports, and use this information to recommend treatment options[33].

Moving forward, the potential for machine learning algorithms to assist doctors is substantial. Tasks such as extracting features from complex data sets like images, ECGs, and other monitoring devices; or spotting patterns indicative of health or illness in individuals from medical records, wearable devices; or combining information from disparate sources to reach diagnoses and treatment decisions, are all well-suited to machine learning approaches. With access to the right kind and volume of training data, machine learning algorithms would be expected to perform well in many of these settings.

Education

In education, machine learning can support new ways of delivering teaching materials, especially in the online environment, and can help teachers to create personalised learning plans for individual students or carry out some routine tasks.

For example, applications are being developed that use machine learning to help teachers to grade student papers more efficiently. One such application – Gradescope – scans students' responses to questions, and groups these according to the answers given. The teacher can then review these groups, checking that the system has allocated students to groups correctly, or manually changing who is allocated to which category. Once the teacher agrees to the proposed groupings, marks can be awarded accordingly. This feedback allows the system to improve its future performance. The system can also automatically compare a student's answer to those of their peers, and direct the student to appropriate resources if they require further study in a particular area[34].

The ability to detect patterns in data and make predictions gives machine learning potential applications in a wide range of fields, including healthcare, education, transport and logistics, public services, finance, pharmaceuticals, energy, the legal sector, manufacturing, and retail.

31. Peng L, Gulshan V. 2016 Deep learning for detection of diabetic eye disease. *Google Research Blog*. See https://research.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html (accessed 22 March 2017).
32. Esteva A *et al*. 2017 Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118. (doi: 10.1038/nature21056)
33. IBM. IBM Watson Health. See http://www.ibm.com/watson/health/oncology (accessed 22 March 2017).
34. See, for example: Abbeel, P. 2016 Machine learning for education (NIPS 2016 workshop). See https://dsp.rice.edu/ml4ed_nips2016 (accessed 22 March 2017).

Machine learning is already used in online education systems; in Massive Open Online Courses (MOOCs) it is used to analyse student inputs, grading tests or other computer-based assignments, and in some computer vision functions. Using machine learning in this way allows course organisers to support a large number of students, and allocate human resources to less routine activities[35].

Further applications are being developed for the classroom, which will be able to track student understanding and make recommendations for future learning activity personalised to individual students to meet their particular needs.

Transport and logistics

To operate safely on the roads, autonomous vehicles need to be able to recognise a range of environmental features, including: obstacles, road signs, pedestrians, and other vehicles. The range and variability of these features means that it is not possible to create hard-coded rules specifying what the vehicle will come into contact with, and how it should respond in different situations. Machine learning allows the vehicle to adapt to a range of features, and respond accordingly. Further developments in the sophistication of autonomous vehicles have applications in a wide range of settings and industries. In one example of this, Amazon is developing the use of delivery drones, with the first successful delivery taking place in December 2016[36].

While autonomous vehicles might be the most high-profile of the potential applications of machine learning in transport, the technology could support a range of functions. To help create intelligent transport systems, for instance, algorithms could analyse historical data on traffic flows in an area, using this data to optimise the system, and to predict how it will respond to different pressures at different times of day. These insights can then be used to reduce congestion, with corresponding implications for reducing carbon emissions. With appropriate information, it would also be possible to assess traffic flows in real time and make dynamic adjustments to improve traffic flow. For example, a network of road sensors, which records vehicle flow and congestion, surrounds the UK's biggest shopping centre in Gateshead. By measuring how vehicles are moving around the centre, it is possible to predict when and where traffic problems will arise, allowing local traffic controllers to intervene before problems occur. Using machine learning, instead of traditional predictive modelling techniques, traffic controllers are able to improve the accuracy of their congestion predictions by up to 50%, which helps ease the strain placed on the local traffic network by shopping centre traffic, as well as reducing emissions and improving shoppers' experiences.

Machine learning can also play a role in optimising logistics and associated processes. This can be through recommending how storage facilities should be set out, so that products can be retrieved most efficiently, or through predicting how much fuel will be required by different delivery vans, based on their likely route and knowledge of traffic flows. Such algorithms are already in successful use in some companies, contributing to improvements in business efficiency and productivity.

35. Hollands F, Tirthali D. 2014 MOOCs: expectations and reality. Columbia University, NY: Center for Benefit-Cost Studies of Education, Teachers College. See http://cbcse.org/wordpress/wp-content/uploads/2014/05/MOOCs_Expectations_and_Reality.pdf (accessed 22 March 2017).

36. Condlifee, J. 2016 An Amazon drone has delivered its first products to a paying customer. *MIT Technology Review*. See https://www.technologyreview.com/s/603141/an-amazon-drone-has-delivered-its-first-products-to-a-paying-customer/ (accessed 22 March 2017).

Public services

For government, machine learning offers the promise of more efficient and effective services, through targeted interventions and tailoring of services. Examples of machine learning and predictive analytics being put to use in tackling public policy issues can already be found, both within the UK and internationally.

Targeting interventions for 'at risk' groups

Some of the most vulnerable individuals in society are often dealing with issues that cut across public services, including housing, health, and justice. These individuals may have multiple points of interaction with the state; identifying these individuals and their needs through advanced data analysis may enable more effective services to be tailored to them. For example, if young people who are at high risk of dropping out of education or failing to find employment can be identified at an early stage, then interventions can be made that seek to prevent these individuals from falling out of the education or training system.

By analysing data from students' school records and related sources, machine learning can be used to create models that predict the likelihood of students becoming NEET (not in education, employment or training) in the future. On the basis of these predictions, schools can intervene at an earlier stage with additional support to encourage those at high risk of becoming NEET to remain in employment or education. Such approaches have been trialled at a local level in the UK, with Essex Council using predictive risk modelling to predict the risk of 14 year olds becoming NEET by age 18[37]. Similar studies have helped Ohio education services to use school records to identify students who are at risk of struggling in school, or failing to graduate[38].

In Kansas, a local jurisdiction is combining datasets across services to analyse the behaviour of service users, looking for patterns that can be used as the basis for predictions about how people will interact with government services. This could help identify those who are in particular need of tailored cross-departmental, coordinated assistance[39] in a way that is also cost effective for government. Such targeted support could help people before a crisis occurs, directing those in need to preventative support in areas such as medical services, mental health services, or community support services. Such analyses require historical data about service use from across departments.

Increasing responsiveness

To effectively manage an incident, such as flooding, government needs to understand what is happening at a local level, predict what might happen next, and decide where to focus efforts accordingly.

Insights relating to each of these can be found in data from first responders, earth observation, or social media. However, there is often limited time available to analyse large quantities of such data, and resources to extract relevant insights might be scarce. Effective static models often already exist, but the ability to run analyses in real time could increase their capabilities.

37. Nuffield Trust. 2011 Predictive modelling for social care: next steps workshop. London, UK: Nuffield Trust.
38. University of Chicago. 2016 Identifying and influencing students at risk of not finishing high school. See https://dssg.uchicago.edu/project/identifying-and-influencing-students-at-risk-of-not-finishing-high-school/ (accessed 22 March 2017).
39. University of Chicago. 2016 Identifying frequent users of multiple public systems for more effective assistance. See https://dssg.uchicago.edu/project/identifying-frequent-users-of-multiple-public-systems-for-more-effective-assistance/ (accessed 22 March 2017).

Machine learning could therefore be put to use in helping to design more effective responses to incidents such as flooding, by combing large datasets to find relevant information, which can be then used as the basis for anticipating how a situation might develop or deciding how resources might be best directed. Machine learning methods could also be used to develop models which take current and past meteorological and environmental data as input and predict changes in flood levels over time.

Finance

Machine learning is already used in banking and finance, for example in systems that detect unusual spending activity, as discussed earlier, or handwriting-recognition systems that allow automated teller machines to read cheques that have been deposited.

Further applications are in development across the sector, including robot bank-tellers that use machine learning to respond to customer queries[40], security systems that use voice recognition to grant customers access to their accounts[41], and there has been speculation that machine learning algorithms could in future help inform monetary policy-making[42].

Pharmaceuticals

The pharmaceuticals sector both relies on and creates large amounts of data, from clinical trials, from drug efficacy studies, or from genetic studies. These large-scale datasets require methods to aid their analysis, in order to extract valuable insights that can improve research and development processes, and to create diagnostic tools to target medicines at patients who will most benefit.

Machine learning could help increase the efficiency of the drug discovery process. For example, machine learning algorithms can analyse molecular structures of potential drug compounds, and predict which of these is likely to be more or less active[43]. Such analysis could help increase the hit rate of screening programmes, thereby identifying more effective drug candidates more quickly.

Another application of machine learning relates to its ability to make predictions, on the basis of patterns in data, about how effective different drugs will be for patients. For example, machine learning has been used to predict how well patients will respond to different drugs used in treating depression[44]. One UK company is using natural language processing to scour published research as a central part of its drug discovery programme[45].

40. McCurry J. 2015 Japanese bank introduces robot workers to deal with customers in branches. *The Guardian*. See https://www.theguardian.com/world/2015/feb/04/japanese-bank-introduces-robot-workers-to-deal-with-customers-in-branches (accessed 22 March 2017).
41. Dunkley E. 2016 Hello, this is your bank speaking: HSBC unveils voice recognition. *Financial Times*. See https://www.ft.com/content/90b635da-d6ea-11e5-8887-98e7feb46f27 (accessed 22 March 2017).
42. Condon, C. 2016 Quest for robo-Yellen advances as computers gain on rate setters. *Bloomberg*. See https://www.bloomberg.com/news/articles/2016-05-24/quest-for-robo-yellen-advances-as-computers-gain-on-rate-setters (accessed 22 March 2017).
43. This area is known as quantitative structure-activity relationships. See, for example: Varnek A, Baskin I. 2012 Machine learning methods for property prediction in chemoinformatics: quo vadis? *J. Chem. Inf. Model*. **52**, 1413–1437. (doi: 10.1021/ci200409x)
44. See, for example: PReDicT (Predicting Response to Depression Treatment). See http://predictproject.p1vitalproducts.com/ (accessed 22 March 2017).
45. The Economist. 2017 Will artificial intelligence help to crack biology? *The Economist*. See http://www.economist.com/news/science-and-technology/21713828-silicon-valley-has-squidgy-worlds-biology-and-disease-its-sights-will (accessed 22 March).

Energy

Machine learning can be used to optimise energy infrastructure. It can analyse patterns of energy use, and use these to design systems that can respond more effectively to peak demands. For example, Google DeepMind has used machine learning to optimise the heating and cooling requirements of its data centres, by predicting temperatures and pressures in the data centre. DeepMind's algorithm was able to reduce the amount of energy needed for cooling by 40%. Having tested the system in a live data centre, DeepMind plans to roll it out more broadly. There may be future applications for a similar system in improving the efficiency of power plants[46].

Legal sector

Recent advances in natural language processing have opened a range of opportunities in the legal sector.

Used alongside chatbot-style interfaces, machine learning can support systems that analyse simple legal queries and provide advice on those queries[47]. The reduced costs associated with such machine learning systems could help increase access to legal services in the market sector that deals with a high volume of lower value cases.

Machine learning can also help carry out routine tasks, such as basic research[48]. It may therefore challenge business models that rely on charging hourly rates, as such routine tasks may be carried out much more quickly.

Machine learning can also be used to predict compliance with the law. For example, it has been used to predict the likelihood that a company or individual will try to evade tax, by learning from patterns of transactions by companies and tax authorities, and using this to simulate how tax evasion schemes are likely to evolve, and which schemes are likely to evade detection in auditing procedures[49]. These opportunities represent different types of disruption to traditional business models, from new ways of providing services to new ways of doing business.

Manufacturing

In manufacturing, machine learning offers an opportunity to automate processes or make them more efficient, create personalised products, or enable predictive maintenance functions. This new narrative for manufacturing in the 21st Century, whereby high-tech manufacturing exploits data-driven technologies and automation, is known as Industry 4.0[50].

46. DeepMind. 2016 Press release: DeepMind AI reduces Google data centre cooling bill by 40%. See https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/ (accessed 22 March 2017).
47. See, for example: Amiquis. See https://amiqus.co/purpose/ (accessed 22 March 2017).
48. See chapter 5.
49. Hemberg E, Rosen J, Warner G, Wijesinghe S, O'Reilly UM. 2015 Tax non-compliance detection using co-evolution of tax evasion risk and audit likelihood. *ICAIL '15*. 79–88. (doi: 10.1145/2746090.2746099)
50. The expression Industrie 4.0 was coined by a German government-sponsored initiative for advanced manufacturing. See, for example: Gartner. 2015 What is Industrie 4.0 and what should CIOs do about it? See http://www.gartner.com/newsroom/id/3054921 (accessed 22 March 2017).

In addition to automating manufacturing processes, machine learning could, for example, change the way in which manufacturing equipment – or manufactured goods – are serviced and maintained. By collecting data about how equipment is operating, and when equipment fails, learning programs can develop predictive maintenance systems. Such systems would anticipate when assets were likely to fail, and direct maintenance work accordingly, thus saving costly repairs at a later date or extended periods of downtime for the failing equipment. For example, predictive maintenance can be used to manage wind turbines, for which downtime can be costly, and regular site inspections may be difficult, owing to their remote location. Being able to accurately estimate when turbines might be at risk of failure – through condition monitoring via sensor data or other patterns – can help deploy staff resources more effectively, while avoiding costly interruptions to services[51].

Advances in machine learning and robotics are also opening new avenues for human-machine interaction in manufacturing. Cobots – robots designed to work in tandem with humans – have, for instance, been used in automotive assembly lines.

Retail

Drawing insight from data about customers, machine learning promises increasingly personalised products, with outputs tailored to the needs or preferences of individual consumers. Machine learning can already make personalised product recommendations, for example recommending potential grocery purchases on the basis of previous shopping history, via recommender systems, discussed earlier in this report (section 1.4).

Further developments in retail could include more intensively-automated shopping experiences, such as those being developed by Amazon, in which shoppers – and their shopping selections – are tracked as they move through a store, and charges are made automatically[52]. Through a combination of sensor-based and machine learning technologies, Amazon Go is able to recognise who is purchasing which products. The result is a functional store with no checkouts.

51. Gauher S. 2016 Evaluating failure prediction models for predictive maintenance. *Cortana Intelligence and Machine Learning Blog*. 19 April 2016. See https://blogs.technet.microsoft.com/machinelearning/2016/04/19/evaluating-failure-prediction-models-for-predictive-maintenance/ (accessed 22 March 2017).

52. BBC. 2016 Amazon unveils plans for grocery shop with no checkouts. *BBC News*. 5 December 2016. See http://www.bbc.co.uk/news/technology-38212818 (accessed 22 March 2017).

2.2 Machine learning in research

By processing the large amounts of data now being generated in fields such as life sciences, particle physics, astronomy, the social sciences, and more, machine learning could be a key enabler for a range of scientific fields, pushing forward the boundaries of science. Machine learning could become a key tool for researchers to analyse these large datasets, detecting previously unforeseen patterns or extracting unexpected insights.

Some early examples of its use in scientific studies are considered below; its potential applications in scientific research range broadly across disciplines, and will include a suite of fields not considered in detail here[53].

New insights in neurosciences

Machine learning has, and has had, a strong influence on modern neuroscience in a number of respects: notably through data analysis[54], and through modelling[55] techniques.

Neuroscience presents considerable data analysis and statistical problems, meaning that supervised, semi-supervised, and unsupervised learning methods are all important tools to enable data analysis across a range of studies[56]. Machine learning can help map how the brain carries out its functions, by finding patterns of activity in vast datasets created by studies of neural activity. By processing images of the brain – such as those created by functional MRI scans – machine learning can correlate areas of activity with specific tasks, such as recognising words[57] or images[58]. The nuances in these images are often too fine to be detected by human analysts, but patterns in them can be discerned by machine learning systems. By providing a deeper understanding of the brain in this way, machine learning could help identify or treat disease in future[59].

Brains face and solve statistically and computationally hard learning problems in order to allow their owners to prosper in noisy, changing, and danger-filled environments. Machine learning offers ways of thinking about and solving these problems in ways that can shed light on equivalent biological solutions, for example by observing a partial match between intermediate representations in deep learning networks and patterns of neural activity in sensory processing pathways in the brain[60].

Machine learning can add new insights to a range of scientific fields.

53. For example, machine learning is expected to have significant applications in the fields of genomics and materials science, amongst others.
54. Kass R, Eden U, Brown E. 2014 Analysis of neural data. Berlin, Germany: Springer Verlag.
55. See, for example: Yamins D, Di Carlo J. 2016 Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci*. **19**, 356–365 (doi: 10.1038/nn.4244), and Gershman S, Niv Y. 2010 Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol*. **20**, 251–256. (doi: 10.1016/j.conb.2010.02.008)
56. Grabska-Barwinska A *et al*. 2017 A probabilistic approach to demixing odors. *Nat. Neurosci*. **20**, 98–106. (doi: 10.1038/nn.4444)
57. Huth A, de Heer W, Griffiths TL, Theunissen FE, Gallant JL. 2016 Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458. (doi: 10.1038/nature17637)
58. Naselaris T, Olman CA, Stansbury DE, Uqurbil K, Gallant JL. 2015 A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage* **105**, 215–228. (doi: 10.1016/j.neuroimage.2014.10.018)
59. Calhoun V, Lawrie SM, Mourao-Miranda J, Stephan KE. 2017 Prediction of individual differences from neuroimaging data. *NeuroImage* **145**, 135. (doi: 10.1016/j.neuroimage.2016.12.012)
60. Graves A *et al*. 2016 Hybrid computing using a neural network with dynamic external memory. *Nature* **538**, 471–476. (doi: 10.1038/nature20101)

Neuroscience has also informed developments in machine learning, for example by inspiring work in the areas of convolutional nets[61], computer vision[62], and episodic memory[63].

Detecting new particles in physics

In July 2012, physicists from the Large Hadron Collider (LHC) at CERN announced that they had discovered the Higgs Boson, an elementary particle which is of critical importance to the Standard Model of particle physics, and which plays a role in giving matter mass.

The Higgs Boson can be created when particles collide together at high energy, as happens in the LHC. Once created, the Higgs Boson quickly breaks down into other particles; it decays within 10^{-22} seconds into other particles, called (gamma) photons[64]. Finding this particle therefore required the detection of a specific pattern of decay amidst the other particle collisions and activity in the LHC.

Machine learning played a role in helping to detect this pattern. Using simulations of what the decay pattern of the Higgs Boson would look like, a machine learning system was trained to pick out this pattern from other activity[65]. Having learned what the presence of the Higgs Boson would look like, the system was put to use on data from the LHC, thereby contributing to the discovery.

Machine learning techniques are used today in many analyses in particle physics, at levels from correctly reconstructing the signals left by individual particles in detectors, and distinguishing these from other particles, to discriminating signals from background noise. These techniques are important in helping to optimise the potential of today's experiments, by increasing the sensitivity of analyses. Typically, at the LHC, they can offer an improvement in sensitivity of between 20% – 40%, meaning that a result which would take two or three years of data to achieve without machine learning can be achieved in substantially less time.

61. LeCun Y, Bengio Y, Hinton G. 2015 Deep learning. *Nature* **521**, 436–444. (doi: 10.1038/nature14539)

62. Fukushima K. 1980 Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern*. **36**, 193–202.

63. Kumaran D, Hassabis D, McClelland J. 2016 What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends Cogn. Sci*. **20**, 512–534. (doi: 10.1016/j.tics.2016.05.004)

64. Brumfiel G. 2012 Higgs triumph opens up field of dreams. *Nature* **487**, 147. (doi: 10.1038/487147a)

65. Castelvecchi D. 2015 Artificial intelligence called in to tackle LHC data deluge. *Nature* **528**, 18. (doi: 10.1038/528018a)

Finding patterns in astronomical data

Research in astronomy generates large amounts of data. For example, once up and running, the Large Synoptic Survey Telescope (LSST), is expected to create over 15 terabytes of astronomical data each night from its images of the night sky[66]. In analysing this data, a key challenge for astronomy is to detect interesting features or signals from the noise, and to assign these to the correct category or phenomenon.

Machine learning can assist in this data analysis, both by preparing the data for use, and by detecting features in the data.

For example, the Kepler mission is seeking to discover Earth-sized planets orbiting other stars, collecting data from observations of the Orion Spur, and beyond, that could indicate the presence of stars or planets. However, not all of this data is useful; it can be distorted by the activity of on-board thrusters, by variations in stellar activity, or other systematic trends. Before the data can be analysed, these so-called instrumental artefacts need to be removed from the system. To help with this, researchers have developed a machine learning system that can identify these artefacts and remove them from the system, cleaning it for later analysis[67].

There is therefore growing interest in using machine learning to analyse the data produced by large survey experiments. The Dark Energy Survey is already using machine learning to estimate photometric redshifts[68], and the next generation of surveys – including the LSST – are preparing to make use of this technology.

Machine learning has also been used to identify new astronomical features, for example:

- Finding new pulsars from existing data sets[69];
- Identifying the properties of stars[70] and supernovae[71]; and
- Correctly classifying galaxies[72].

66. LSST (Large Synoptic Survey Telescope). See https://www.lsst.org/ (accessed 22 March 2017).
67. Roberts S, McQuillan A, Reece S, Aigrain S. 2013 Astrophysically robust systematics removal using variational inference: application to the first month of Kepler data. *Mon. Not. R. Astron. Soc.* **435**, 3639–3653.
68. Sadeh I, Abdalla F, Lahav O. 2016 ANNz2: photometric redshift and probability distribution function estimation using machine learning. *Publ. Astron. Soc. Pac.* **128**, 104502. (doi: 10.1088/1538-3873/128/968/104502)
69. Morello V, Barr ED, Bailes M, Flynn CM, Keane EF, van Straten W. 2014 SPINN: a straightforward machine learning solution to the pulsar candidate selection problem. *Mon. Not. R. Astron. Soc.* **443**, 1651–1662. (doi: 10.1093/mnras/stu1188)
70. Miller A *et al.* 2015 A machine learning method to infer fundamental stellar parameters from photometric light curves. *Astrophys. J.* **798**, 17. (doi: 10.1088/0004-637X/798/2/122)
71. Lochner M, McEwen JD, Peiris HV, Lahav O, Winter MK. 2016 Photometric supernova classification with machine learning. *Astrophys. J.* Suppl. Ser. **225**, 31. (doi: 10.3847/0067-0049/225/2/31)
72. Banerji M *et al.* 2010 Galaxy Zoo: reproducing galaxy morphologies via machine learning. *Mon. Not. R. Astron. Soc.* **406**, 342–353. (doi: 10.1111/j.1365-2966.2010.16713.x)

Understanding the effects of climate change on cities and regions

The current generation of climate models can be used to make global predictions under differing scenarios of future climate change, for example changes to global temperature, precipitation patterns, or sea level. Using the known properties of the physical systems involved, these models can accurately predict the broad-scale consequences of climate change, but shed less light on future climate change on regional scales.

To inform decision-making at a national or local level, these predictions need to be understood in terms of their consequences for cities or regions; for example, predicting the number of summer days where temperatures exceed 30°C within a city in 20 years' time[73]. These areas might have access to detailed observational data about local environmental conditions – from weather stations, for example – but it is difficult to create accurate projections from these alone, given the baseline changes taking place as a result of climate change.

Machine learning can help bridge the gap between these two types of information. It can integrate the low-resolution outputs of climate models with detailed, but local, observational data; the resulting hybrid analysis would improve the climate models created by traditional methods of analysis, and provide a more detailed picture of the local impacts of climate change.

For example, a current research project at the University of Cambridge[74] is seeking to understand how climate variability in Egypt is likely to change over coming decades, and the impact these changes will have on cotton production in the region. The resulting predictions can then be used to provide strategies for building climate resilience that will decrease the impact of climate change on agriculture in the region.

73. Banerjee A, Monteleoni C. 2014 Climate change: challenges for machine learning (NIPS tutorial). See https://www.microsoft.com/en-us/research/video/tutorial-climate-change-challenges-for-machine-learning/ (accessed 22 March 2017).

74. See ongoing work at the British Antarctic Survey on machine learning techniques for climate projection.

2.3 Increasing the UK's absorptive capacity for machine learning

The wide range of everyday machine learning examples outlined above gives a taste of the potential economic and social benefits associated with machine learning.

In practice, the benefits of machine learning will be delivered in different ways in different applications. Achieving the benefits of machine learning will rely on increasing the absorptive capacity of UK industry, so that it can make use of machine learning. With this in mind, building on 2015's *Transforming our Future* conference[75], the Royal Society convened senior representatives from a range of fields – manufacturing, pharmaceuticals, energy, cities, transport, and the legal sector – to consider how their sectors might best enjoy the potential benefits of machine learning. Insights from this work inform the recommendations for action across the following chapters.

The need for an amenable data environment, strong skills pipeline, support for business, and a governance system that engenders confidence in the applications of machine learning are consistent across sectors and applications. The chapters that follow make recommendations to help increase the absorptive capacity for machine learning in the UK, and ensure that the benefits of machine learning are broadly shared, through action in each of these areas.

75. The Royal Society. 2015 Machine learning conference report. 22 May 2015.

Chapter three
Extracting value from data

Left
Machine learning is a key tool for making sense of ‘big data’, and creating value from it. © Pobytov.

Extracting value from data

One way of thinking about many (but not all) applications of machine learning is that the algorithms learn from examples – called training data. The recent success of machine learning in matching or even improving on human behaviour for some tasks, including image and speech analysis, is due in no small part to the recent explosion of training data in these areas: the vast number of examples on which to train the algorithms has been a critical part of their improved performance.

Machine learning is therefore both a method that requires data and a tool that enables uses of it; access to data is required to create machine learning methods and train machine learning systems, and these systems can be put to use in making sense of the large, and growing, amount of data available today. In extracting valuable information from data, machine learning can help realise the social and economic benefits expected from so-called 'big data'.

In turn, machine learning requires a 'machine-friendly' data environment, based on open standards that make using open data easier. This chapter assesses some of the issues around data availability for machine learning.

3.1 Machine learning helps extract value from 'big data'

Ninety percent of the world's data has been created within the last five years[76]. In this age of 'big data', an increasing volume of information is being collected, from a greater range of sources, and at greater speed than ever before. Image or video uploads to social media, GPS-enabled devices, and other online activities are generating stores of data, as people spend more of their work and leisure time online. This all contributes to the creation of an estimated 2.5 billion gigabytes of data per day[77].

These changes to the volume, variety, and velocity[78] of data collection have created a potentially rich resource for the digital economy. One estimate suggests that open data could help create $3 trillion of value each year for the global economy[79]. Early indicators of economic growth show promise in this area; digital industries grew 32% faster than the rest of the UK economy from 2010 to 2014, with employment in these sectors growing 2.8 times faster than in other sectors of the economy[80].

In this new data environment, considerable economic benefits are therefore at stake: data has been described as the 'new oil' for the digital economy; a resource with the potential to support a new industrial revolution.

76. The Royal Society. Learning infographic: what is machine learning? See https://royalsociety.org/topics-policy/projects/machine-learning/machine-learning-infographic/ (accessed 22 March 2017).

77. Wall W. 2014 Big data: are you ready for blast off? *BBC News*. 4 March 2014. See http://www.bbc.co.uk/news/business-26383058 (accessed 22 March 2017).

78. Executive Office of the President. 2014 Big data: seizing opportunities, preserving values. Washington, US: The White House. 1 May 2014. See https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf (accessed 22 March 2017).

79. Manyika J, Chui M, Farrell D, van Kuiken S, Groves P, Doshi E. 2013 Open data: unlocking innovation and performance with liquid information. McKinsey Global Institute. See http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information (accessed 22 March 2017).

80. The Royal Society. 2016 Progress and research in cybersecurity. See https://royalsociety.org/topics-policy/projects/cybersecurity-research/ (accessed 22 March 2017).

The nature of this new data environment challenges traditional analytical approaches in data science and statistics, as the complexity of big data can limit the effectiveness of existing methods of analysis. The complexity and scale of data available today therefore demands new tools to create valuable insights. In this context, machine learning has a vital role to play in making sense of large quantities of potentially dynamic data. It can process volumes of data that would be unmanageable for humans, picking out the patterns that subsequently become useful. Machine learning therefore extracts value by deriving new insights from the mass of data, and in turn data is needed to develop machine learning, by training systems to detect patterns or make predictions.

As noted above, data has been described as the new oil; holding incredible economic potential, but requiring refinement in order to realise this. If not the new oil in and of itself, then data is at least the fuel for machine learning, and a data environment that enables the effective use of data will be key to enabling machine learning to be put to use, and hence to deliver its promised benefits.

3.2 Creating a data environment to support machine learning

Data now comes from a range of sectors – individual, public, private, non-profit, and academic – and in a range of formats. Each of these data sources comes with specific challenges, and the diversity of sources requires new approaches to managing data in a machine-friendly way. Open data is defined as "data that is published under a licence with express permission to re-use, share and modify[81]". Some of this data is public, but not all public data is open, nor does it need to be. Data can be accessible without being usable, for example owing to practical considerations relating to its quality. Conversely, data which is not open can be made accessible to certain users via framework or access agreements.

Data availability across sectors is important

Public sector data can be a key enabler

Access to public sector data could catalyse a range of economic activity: the direct value of public sector information to the UK economy has been estimated at £1.8 billion[82], with wider social and economic benefits from this totalling up to £6.8 billion[83].

There are different sorts of data held in the public sector. Some of this is social[84], some is not directly related to individuals[85], while some (for example in the NHS) relates to confidential personal information. Healthcare and related data raises particular issues which are addressed later.

The UK Government has already supported a series of initiatives to make public sector data accessible, including providing funding for the Open Data Institute, data.gov.uk, and the Administrative Data Research Network. This support has driven valuable progress in ensuring the openness of public data, and the UK ranks

81. ODI. 2015 Open data roadmap for the UK – 2015. See http://theodi.org/roadmap-uk-2015 (accessed 22 March 2017).
82. Deloitte. 2013 Market assessment of public sector information: report commissioned for the Department for Business, Innovation and Skills. See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf (accessed 22 March 2017).
83. Shakespeare S. 2013 An independent review of public sector information. See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf (accessed 22 March 2017).
84. Information shared by individuals, for example location or shared links.
85. For example, economic data.

well in international measures of open data[86]. Efforts to increase the availability of government data have already resulted in over 40,000 datasets being made open via data.gov.uk[87].

The UK has already committed to the G7 Open Data Charter, the principles of which state that government data should be openly published by default, usable by all, and released in high-quality formats that allow automated processing[88]. Building on this, Government has identified the need to create "a high quality national information infrastructure" that will facilitate access to data[89], reiterating its commitment to opening up data via gov.uk in 2017's UK Digital Strategy[90]. Continuing to make data sources available in this way, where appropriate, can support further developments in machine learning.

In addition to making further progress with open data efforts, it is necessary to recognise the value of some public sector data. While making such data open can bring benefits, considering how those benefits are distributed is important. As machine learning becomes a more significant force, the ability to access data becomes more important, and those with access can attain a 'first mover feedback' advantage that can be significant. When there is such value at stake, it will be increasingly necessary to manage significant datasets or data sources strategically.

One of the UK's key data assets lies in the NHS. NHS data is a unique information resource for the UK, and making effective use of it will be vital for 'UK plc'. Given the nature of this data, there are natural and legitimate concerns about how the value of its use can be balanced against concerns about privacy, and questions about who should be able to access it. Such tension between public good and personal privacy form a core part of the Royal Society's and British Academy's work on data governance[91].

Any access to NHS data must clearly be under very carefully regulated conditions to ensure protection of individual privacy. There may be helpful models here in those developed for access to healthcare information in biomedical research studies, for example the UK Biobank[92] data access agreements, or frameworks used by the Wellcome Trust Case Control Consortium[93]. If this balancing act is resolved, and if appropriately controlled access mechanisms can be developed, then there is huge potential for NHS data to be used in ways that will both improve the functioning of the NHS and improve healthcare delivery.

86. World Wide Web Foundation. 2016 Open Data Barometer – 3rd Edition. See http://webfoundation.org/about/research/open-data-barometer-3rd-edition/ (accessed 22 March 2017).
87. DCMS (Department for Culture, Media & Sport). 2017 Policy paper: UK digital strategy. See https://www.gov.uk/government/publications/uk-digital-strategy (accessed 22 March 2017).
88. Cabinet Office. 2013 Policy paper: G8 open data charter. See https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex (accessed 22 March 2017).
89. Cabinet Office. 2016 Policy paper: UK open government national action plan 2016-18. See https://www.gov.uk/government/publications/uk-open-government-national-action-plan-2016-18/uk-open-government-national-action-plan-2016-18 (accessed 22 March 2017).
90. DCMS (Department for Culture, Media & Sport). 2017 Policy paper: UK digital strategy. See https://www.gov.uk/government/publications/uk-digital-strategy (accessed 22 March 2017).
91. Acceptable uses of NHS data have also been considered in a range of work by the Wellcome Trust and the National Data Guardian. See, for example: National Data Guardian. 2016 Review of data security, consent, and opt-outs. See https://www.gov.uk/government/publications/review-of-data-security-consent-and-opt-outs (accessed 22 March 2017).
92. UK Biobank. See http://www.ukbiobank.ac.uk/ (accessed 22 March 2017).
93. Wellcome Trust Case Control Consortium. See https://www.wtccc.org.uk/ (accessed 22 March 2017).

RECOMMENDATIONS

Good progress in increasing the accessibility of public sector data has positioned the UK as a leader in this area; continued efforts are needed in a new wave of 'open data for machine learning' by Government to enhance the availability and usability of public sector data, while recognising the value of strategic datasets.

In areas where there are datasets unsuitable for general release, further progress in supporting access to public sector data could be driven by creating policy frameworks or agreements which make data available to specific users under clear and binding legal constraints to safeguard its use, and set out acceptable uses. The UK Biobank demonstrates how such a framework can work. Government should further consider the form and function of such new models of data sharing.

Access to proprietary data

The public sector is not the only originator of rich data resources; a lot of data is held in private companies, and there is significant economic value which could be unlocked through its use. It is estimated that £66 billion of business and innovation opportunities could be generated through effective use of data[94] in the UK[95]. Some of this data will relate directly to the company's business. For some digital companies, there will also be data on individual preferences and behaviour collected while people interact with the company's app.

In discussing the openness of private sector – or proprietary – data, there is a clear tension between the need to maintain both proprietary advantage and the privacy of service users, and the potential advances that could be made via machine learning if such data were made available. Furthermore, anonymisation of data can be time consuming and costly, difficult to guarantee if data is linked, and other legal barriers may discourage data sharing.

If such data is sensitive, or offers significant competitive advantage, then clearly it makes little sense to share this between companies. Not all data, however, has such value, and there may be benefits to sharing data within or between companies. There may also be settings in which it is beneficial for companies to share summaries of their data with their competitors, for mutual benefit.

94. DCMS (Department for Culture, Media & Sport). 2017 Policy paper: UK digital strategy. See https://www.gov.uk/government/publications/uk-digital-strategy (accessed 22 March 2017).

95. Parris S, Spisak A, Lepetit L, Marjanovic S, Gunashekar S, Jones M. 2015 The Digital Catapult and productivity: a framework for productivity growth from sharing closed data. RAND Research Report RR-1284-DC. See http://www.rand.org/content/dam/rand/pubs/research_reports/RR1200/RR1284/RAND_RR1284.pdf (accessed 22 March 2017).

For example, data can be used in benchmarking or in collaboratively creating new information. To encourage such behaviour, businesses need to see the potential benefits of sharing data in this way, and to see lower barriers to entry, via platforms or marketplaces which facilitate data sharing.

Privacy-preserving machine learning systems, and other mechanisms to support access to datasets, are an active area of research, where further work could generate solutions to some of these issues (see chapter 6).

Data from research

The collection of experimental and observational data has always been at the heart of the scientific endeavour.

As computing power has increased, and new technological capabilities have increased the scale of data collection in research, the volume and variety of data available to researchers has also increased. In some areas of science (such as astronomy, particle physics, and genomics) the volumes of data routinely generated in scientific studies are huge. For example, the Square Kilometre Array – a powerful new telescope that will be used to survey the night sky – has the potential to generate more data each second than the internet itself does[96]. As noted earlier, machine learning techniques can be directly helpful to the researchers who generate the data, in their efforts to analyse and interpret it. But research datasets can also have value to others, beyond their initial analysis, if they are available.

There are very strong arguments – related to the openness, transparency, and reproducibility of research – for research data to be made available to others, not least to allow checking of the original analyses and conclusions. Journals are increasingly insisting on data being made available as a condition of publication. Research funders are also increasingly insisting on data being made available to leverage the cost of the original data collection by allowing other scientists to work on the data, thereby increasing the potential scientific discovery from it. In some cases there will also be commercial opportunities in analyses of such data, including through the use of machine learning techniques.

Where data concerns individuals, for example in biomedical research, additional issues arise. Individuals in research studies will have given explicit informed consent and this consent will also include details of uses to which their data may be put. These consents can restrict further use of the data, for example by saying that it will only be used to study a particular disease, or will not be provided to commercial organisations. As awareness has risen of the extent to which there is scientific value in combining data from different studies, there has been a general trend, which should be encouraged, to make consents broader where this is ethically acceptable. The distinction between 'commercial' and 'non-commercial' use of data is also an increasingly difficult one to draw: academic groups commonly collaborate with colleagues in the commercial sector, or having carried out academic research, which generates intellectual property, can licence this to commercial organisations or create spin-out companies to exploit it.

96. The Royal Society. 2015 Response to the House of Commons Science and Technology Committee's inquiry into the big data dilemma. 3 September 2015. See https://royalsociety.org/topics-policy/publications/2015/big-data-dilemma/ (accessed 22 March 2017).

There are now good working models for making available biomedical research data on individuals to *bone fide* researchers for use consistent with the participant consents and with appropriate safeguards to protect individual privacy and confidentiality[97]. Where existing consents limit the extent to which individual level data can be made available, making available summaries of the data, which are consistent with the consent given, would allow others to perform analyses of the data, and frameworks to enable this should be encouraged. The practice of making available summary statistics from large genetic studies of human disease has already become fairly widespread, and is often insisted on by journals in this field. These summary statistics can be used to carry out most of the types of analysis that scientists not directly involved in a study would be interested in, thus widening the potential uses of the data, while protecting the privacy of individuals in the study.

For the benefits of data availability to be fully realised, data from research needs to be produced in a way that makes it: accessible, so others can find it; intelligible, so it can be scrutinised; assessable, so its reliability can be judged before use; and usable by others[98].

Publishing data in this way can also help increase the impact of research. For example, in a study of cancer microarray data, the co-publication of publicly available data was found to be associated with a 69% increase in citation of the original publication[99].

As data management and data availability become an ever-more integral part of science, the need to bring in specific expertise in handling or processing data, and in preparing it for release, will have implications for the allocation of research funding. While resource costs, such as staff costs, can be considered as part of funding applications to research councils, guidance on the extent to which applications for funding may cover data handling is not clear; while some schemes may offer this, it is not clear that this is always the case.

There are many advantages to openness in academic data, as noted in the Royal Society's report on *Science as an open enterprise*. Approaches to, and culture surrounding, openness of research data vary across academic fields, and funders have a key role in helping to shape these.

97. Notable examples of these practices include the UK Biobank (See http://www.ukbiobank.ac.uk/) and the Wellcome Trust Case Control Consortium (See https://www.wtccc.org.uk/).

98. The Royal Society. 2012 Science as an open enterprise. See https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/ (accessed 22 March 2017).

99. This increase was independent of journal impact factor, publication data, or author country of original. See: Piwowar H, Day R, Fridsma D. 2007 Sharing detailed data is associated with increased citation rate. *PLoS ONE*, **2**, e308.

RECOMMENDATIONS

Continuing to ensure that data generated by charity- and publicly-funded research is open by default and curated in a way that facilitates machine-driven analysis will be critical in supporting wider use of research data. Where appropriate, journals should insist on this data being made available to other researchers in its original form, or via appropriate summary statistics where sensitive personal information is involved.

In designing their studies, researchers should consider future potential uses of their data, and build in the broadest consents that are ethically acceptable, and acceptable to research participants.

Research funders should ensure that data handling, including the cost of preparing data and metadata, and associated costs, such as staff, is supported as a key part of research funding, and that researchers are actively encouraged across subject areas to apply for funds to cover this. Research funders should ensure that reviewers and panels assessing grants appreciate the value of such data management.

3.3 Extending the lifecycle of open data requires open standards

Machine learning enables new uses of open data, processing large datasets or combining data in innovative ways to extract new insights. These new uses potentially extend the lifecycle of data. During this lifecycle, datasets need to remain meaningful to a range of different users. These include, for example:

- data scientists who need to understand what the data represents, how it was created, its context and how it should be used – for example in terms of its registry catalogue or licensing requirements;
- data-driven service users who may wish to understand what data has been used to develop algorithms, the contexts for which the behaviour of the algorithm has been tested and where the data comes from;
- compliance functions which need to understand the 'journey' taken by the data and whether it is being used in an approved way; and
- data owners who may wish to check that if they are selling or contributing data, their data is being used in an approved way.

This broader lifecycle and increased number of users increases the importance of understanding the history of a dataset, its provenance, the context in which it was created or is used, its meaning, and its quality.

Open data, and data made available through appropriate release mechanisms, can be messy or erroneous, with incorrectly labelled fields, or gaps in individual entries: feeding such data into machine learning systems risks creating erroneous results. Preparing data for use by machine learning algorithms can take considerable effort, and processing messy data so that it is suitable for analysis can occupy a significant proportion of the time spent on a data mining project. Therefore making data open or available is necessary, but is not alone sufficient for enabling machine-led analysis. Data needs to be both accessible, and machine-ready, or otherwise curated.

This process of curation involves transforming data into usable forms, for example by:

- reducing errors or inconsistencies within data as a result of inaccuracies in collection;
- combining the data with metadata to ensure its characteristics are accurately recorded;
- making sure the provenance of the data is clear, so the user can understand its characteristics and any restrictions on its use; and
- integrating complex or heterogeneous data sources, for example data at different resolutions, to make these compatible with each other before analysis.

Some of this work can be automated, as noted later in this report (see chapter 6). Until these technical solutions are more widespread, the adoption of clear and open standards can make this process – and therefore the use of data – easier.

An understanding of what type of data is being used, how it is processed, and under what conditions, is now needed throughout the lifecycle of data, so that each user is able to understand the provenance of the data they are using, and the significance of this for their analyses.

Such information is encoded via metadata: information associated with a dataset that tells a user where the data is located, how it is structured, what it means, and whether there are restrictions on its use. Metadata helps define the meaning of the data and describes its provenance. In an environment where more people are using more data for different purposes, the standards associated with metadata need to be meaningful to a range of users.

Standards can help make sure that the meaning of data is retained, as it is transferred between systems, by setting out where the data came from and how it has been processed. There are currently many different standards for metadata, with thousands of technical standards around data and metadata each covering a tiny fragment of the problem, which vary in their requirements and in the extent to which they are implemented. With so many standards to choose from, implementation becomes patchy. This patchwork of use creates islands of data use, where specific standards are used for specific reporting requirements, resulting in reduced interoperability between systems.

Open standards are intended to make the exchange of data easier, by discouraging the creation of specific systems that tie data – or consumers – to individual providers, and by allowing software systems, datasets, or documents to be interoperable, and put to a wide range of uses. In doing so, they can open the field to new providers.

RECOMMENDATION

New open standards are needed for data, which reflect the needs of machine-driven analytical approaches.

The Government has a key role to play in the creation of new open standards, for example for metadata. Government should explore ways of catalysing the safe and rapid delivery of these to support machine learning in the UK.

3.4 Technical alternatives to open data: simulations and synthetic data

For some applications of machine learning or in developing some machine learning methods, simulated or synthetic data can be used as an effective, or desirable, alternative to 'real-world' datasets.

Simulations are an imitation of a 'real-world' environment or task; using some observed data, they model the environment or task at hand. In the process, simulations can generate a large amount of data, potentially associated with different environmental states. This data can be used to train machine learning systems, without relying on external sources.

For applications where paucity of data, rather than its absence, is a limiting factor in the use of machine learning, simulations can be used to generate data to train a system. For example, when training AlphaGo to beat human players at the ancient Chinese game of Go (see Box 1), researchers at Google DeepMind initially trained the system using a dataset comprised of 30 million moves from online games played by human amateur players. This enabled the system to predict human moves in a given position with 57% accuracy[100]. To strengthen its abilities – and ultimately enable it to beat the best human professional players – the system then played thousands of simulated games against itself, incrementally learning from its mistakes. Ultimately millions of these self-play games – the synthetic data – were used to generate an evaluation function, which AlphaGo used to predict which side was winning in a particular position, and by how much, a capability critical to making good decisions about what move to make next.

100. Silver D, Hassabis D. 2016 AlphaGo: Mastering the ancient game of Go with machine learning. *Google Research Blog*. See https://research.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html (accessed 22 March 2017).

The ability to test multiple predictions and outcomes before committing to a course of action makes simulations particularly useful in the development of machine learning systems where it is difficult to safely or meaningfully test the outcomes of predictions in the real world. In these cases, the issue at hand is not simply access to data, but the difficulty of testing the outcomes of machine learning models in a safe and efficient way. As a result, simulations can be particularly useful in applications where the cost of failure is high, or there is little scope for real-world testing of multiple outcomes.

For instance, many applications of robotics require autonomous systems to navigate their environment, without damaging themselves or the objects with which they interact. For such systems, an accidental collision or fall could cause damage to the hardware that would be expensive and time consuming to correct. Yet if these systems are too cautious in their movements, or require specific routes to be manually programmed, then their overall usefulness is diminished. One approach to addressing these difficulties is for the system to run a simulation of its environment, computing the different moves it could make and the consequences of those actions, before committing to the physical act of movement. The simulated environment enables the system to explore a wide range of potential behaviours, and to calculate which would be the most effective[101]. This type of approach is already being used to help teach robots how to walk[102].

Further applications might include models of economic or physical systems, where there is little or no opportunity to test the accuracy or implications of different predictions or decisions.

For example, in weather forecasting, machine learning-based simulations can investigate a range of potential atmospheric dynamics, and use the most accurate results from these simulations to help predict the weather. The coupling of these with deep learning methods can increase the accuracy of results[103].

Simulations can be created algorithmically, using a machine learning method known as generative models. Using features in data they have been exposed to, these models are able to generate similar data, on the basis of having learned the features of the dataset. The sophistication and application of simulations and synthetic data could therefore be further advanced through the use and development of generative models.

101. See, for example: Mordatch I, Mishra N, Eppner C, Abbeel P. 2016 Combining model-based policy search with online model learning for control of physical humanoids. IEEE International Conference on Robotics and Automation (ICRA). (doi:10.1109/ICRA.2016.7487140)

102. Knight W. 2015 Robot toddler learns to stand by "imagining" how to do it. *MIT Technology Review*. 6 November 2015. See https://www.technologyreview.com/s/542921/robot-toddler-learns-to-stand-by-imagining-how-to-do-it/ (accessed 22 March 2017).

103. Grover A, Kapoor A, Horvitz E. 2015 A deep hybrid model for weather forecasting. *Proceedings of the 21st ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*, 379–386. (doi:10.1145/2783258.2783275)

Creating good simulations of real-world data is often a hard problem, which requires a thorough understanding of the complexities of the system giving rise to the data under consideration. Use of simulations therefore requires some care. If a simulation fails to capture important properties of the real-world problem being investigated, analyses based on its data can be misleading. Even if a simulation is highly accurate in contained, well-defined environments, and the resulting model is trained on the basis of accurate data from this environment, complexity or nuance from a real world scenario may be lost, resulting in decreased accuracy or incorrect results when the model is deployed in the real world. As a consequence, considerable caution is needed before relying too heavily on simulated data in many real-world settings.

Chapter four
Creating value from machine learning

Left
Building skills at all levels – from data literacy to advanced machine learning – will be important to bring about the benefits of machine learning. This includes introducing key concepts in schools.

Creating value from machine learning

4.1 Human capital, and building skills at every level

Much has already been written about the need to build the UK's digital skills base, noting the persistent shortage of people with digital skills for digital jobs in the UK[104]. Almost a quarter of the UK's population lack basic digital skills, let alone an understanding of machine learning[105]. Furthermore, in England, Wales and Northern Ireland, fewer than one in five students study any maths after age 16[106].

Meanwhile, top end estimates suggest that 58,000 new data science jobs are created each year, and in many sectors it is clear that demand for skills in data science and related fields is outstripping supply. GSK, for example, has reported that 25% of their vacancies in data science have remained vacant over the past 18 months[107].

To thrive in an environment augmented by machine learning, and in which machine learning is a key tool for daily activities and work, citizens will require data literacy skills, which enable them to use and interact with data, and an understanding of the strengths and weaknesses of technologies such as machine learning.

In addition to a general need for everyone to have some basic level of digital literacy in a world where machine learning is more widespread, there are more-specific and more-specialised skills gaps that need addressing. One of these is for what might be called 'informed-users' of machine learning algorithms. These will be individuals working in particular commercial areas who interact directly with machine learning algorithms in order to add value for their organisations. They will need the knowledge and expertise to make informed choices about which existing algorithms to use and what the strengths and weaknesses of these are.

To date, the UK has played a leading role in the development of machine learning algorithms at the cutting edge of research in the field. The demand and opportunities here are likely to grow, so for the UK to maintain its strong position there is another skills need that requires attention, namely producing the next generation of researchers and research leaders in machine learning.

These three skills needs will now be addressed in turn.

104. See, for example: Ecorys. 2016 Digital skills for the UK economy: research paper commissioned for the Department for Business, Innovation & Skills, and Department for Culture, Media & Sport. See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/492889/DCMSDigitalSkillsReportJan2016.pdf (accessed 22 March 2017).

105. Shadbolt N. 2013 Shadbolt Review of computer sciences degree accreditation and graduate employability. See http://dera.ioe.ac.uk/16232/2/ind-16-5-shadbolt-review-computer-science-graduate-employability_Redacted.pdf (accessed 22 March 2017).

106. The forthcoming Smith Review will investigate these issues in more detail. For example, see: Hodgen J, Pepper D, Sturman L, Ruddock G. 2010 Is the UK an outlier? An international comparison of upper secondary mathematics education. London, UK: Nuffield Foundation. See http://www.nuffieldfoundation.org/sites/default/files/files/Is%20the%20UK%20an%20Outlier_Nuffield%20Foundation_v_FINAL.pdf (accessed 22 March 2017).

107. The Royal Society. 2016 Data analytics: the skills need in STEM. Conference report. See https://royalsociety.org/~/media/events/2016/11/data-science-workshop/data-analytics-conference-report-16112016.pdf (accessed 22 March 2017).

BOX 2

Introducing key concepts in machine learning

Key Stage 2 (Age 7)

What a computer program is, and how it works[108].

How algorithms sort datasets[109].

What intelligence means, the different types of intelligence, and how this relates to computer programming[110].

How computer systems can learn[111].

Simple examples of how machine learning works[112].

Ethical considerations in automated decision-making[113].

Key Stage 5 (Age 18)

The consequences of false positive or false negative decisions[114].

A basic grounding in what machine learning is, and what it does, will be necessary to interact with these systems as they become more pervasive.

General skills for a world of widespread machine learning

As daily interactions with machine learning – either at work or at leisure – become the norm, a basic grounding in what machine learning is, and what it does, will be necessary in order to grasp, at a basic level, how our data is being used, and what this means for the information presented to us by machine learning systems. Such digital citizenship will become an essential tool for navigating the digital world, and will need to be learned by people of all ages.

If introduced at primary or secondary school, a basic understanding of key concepts in machine learning can help with navigating this world, and encourage further uptake of data science subjects. Box 2 sets out areas in which machine learning can be introduced to teaching from Key Stage 2 upwards.

108. See, for example: McOwan P, Curzon P. 2008 The intelligent piece of paper. See http://www.cs4fn.org/teachers/activities/intelligentpaper/intelligentpaper.pdf (accessed 22 March 2017).
109. See, for example: Oxford Sparks. 2017 Key Stage 3 – All sorted! See http://www.oxfordsparks.ox.ac.uk/content/teaching-resources (accessed 22 March 2017).
110. See, for example: McOwan P, Curzon P. 2008 The intelligent piece of paper. See http://www.cs4fn.org/teachers/activities/intelligentpaper/intelligentpaper.pdf (accessed 22 March 2017).
111. See, for example: Curzon P, McOwan P, Black J. 2008 Artificial intelligence, but where is the intelligence? See http://www.cs4fn.org/ai/downloads/aiwhereistheintelligence.pdf (accessed 22 March 2017).
112. See, for example: Curzon P, McOwan P, Black J. 2008 Artificial intelligence, but where is the intelligence? See http://www.cs4fn.org/ai/downloads/aiwhereistheintelligence.pdf (accessed 22 March 2017).
113. See, for example: Oxford Sparks. 2017 Key Stage 4 – Computer says no. See http://www.oxfordsparks.ox.ac.uk/content/teaching-resources (accessed 22 March 2017).
114. See, for example: Oxford Sparks. 2017 Key Stage 5 – Testing testing. See http://www.oxfordsparks.ox.ac.uk/content/teaching-resources (accessed 22 March 2017).

A range of school activities to encourage data literacy already exist, and these can help build understanding of machine learning.

BOX 3

Initiatives to increase data literacy via schools or enrichment activities

- Science, Technology, Engineering, and Mathematics (STEM) Clubs give school students the chance to explore aspects of science, technology, engineering and maths. STEM Clubs tend not to be curriculum based, but can focus on specific disciplines or go across STEM subjects. A STEM Club could be about computer coding, maths puzzles, robotics, engineering, chemistry, astronomy, or another related subject. The National STEM Clubs Programme is funded by the Department for Education, the Gatsby Charitable Foundation and the Scottish Government.
- STEM Ambassadors get involved in a huge range of activities, which can all have an impact on young people's learning and enjoyment of STEM, including: giving careers talks or helping at careers fairs; providing technical advice or practical support to STEM projects in the classroom; supporting projects in after-school STEM Clubs; judging school STEM competitions; speed networking with pupils, parents and teachers; devising or delivering practical STEM experiments or demonstrations; or helping students with mock job interviews.
- Code Clubs: There are 4,892 Code Clubs in the UK, teaching over 68,000 learners. These provide support for a nationwide network of volunteers and educators who run free coding clubs for children aged 9 – 11. Code Club was founded in 2012, and in 2015 joined forces with the Raspberry Pi Foundation, a registered UK charity.

In 2012, the Society reviewed school Information Communication Technology curricula in its report *Shut down or restart*. Following the report's publication, many positive changes have taken place to school computing curricula in the UK. In England, the new school computing curriculum provides a prime opportunity to embed these concepts from age 5. The Society is in the process of reviewing the progress made towards the aspirations set out in its report. Keeping fast-changing disciplines relevant presents challenges for teachers and schools. The Computing Education Advisory Group is exploring how to harness expertise from businesses and academia to support schools, many of which do not have access to expert computing teachers. The rapidly evolving data science needs of other disciplines will need to be considered in future curricula and qualification reviews.

In addition to these school-based measures, there are already many extra-curricular initiatives to encourage data literacy, uptake of mathematical or computer sciences, and further – or better – study of related fields (see Box 3). Rather than compete with, or be segregated from, these initiatives, machine learning should be factored into both curricular and enrichment activities in this space, and in relation to scientific and other non-scientific subjects[115, 116].

Steps to increase machine learning literacy from a young age may therefore include:

- Key concepts in machine learning being introduced as part of the computing curriculum, with students interacting with or coding machine learning algorithms in practical classes (see Box 2, for examples of where and how these concepts can be taught).
- Insights from machine learning – or examples of how these systems work – being used in science classes or in non-scientific disciplines.
- Discussions about key ethical concepts in machine learning, and the governance of access to personal data, in ethics classes.
- Code Club modules which enrich students' understanding of machine learning by showing how machine learning algorithms are put together.

Across these areas, activities should be designed in a way that ensures they are inclusive and appealing to a broad range of students, for example to help address the underrepresentation of women in machine learning-related careers.

115. One example of this could be image recognition in biological sciences, the applications of which were detailed in earlier chapters. Example resources to teach related concepts include: Oxford Sparks. 2017 Key Stage 4 – Picture this. See http://www.oxfordsparks.ox.ac.uk/content/teaching-resources (accessed 22 March 2017).

116. For example, ethics classes.

A pipeline of informed users or practitioners is also needed to help put machine learning to use.

As the Royal Society's *Vision for science and mathematics education* report noted, "the ability of people to understand the world in which they live and work increasingly depends on their understanding of scientific ideas and associated technologies and social questions", and, for most people, this understanding will be achieved through education[117]. Achieving this understanding can help citizens to participate more fully in the democratic process, in addition to enhancing research and the economy[118].

Machine learning therefore represents part of a wider need to ensure there is a good grounding of STEM for future citizens, and that people have the best possible basis for dealing with potential future career changes during their working lives, by developing broad skills. With this in mind, a broad and balanced curriculum that includes science and maths education up to age 18 could support a broad skills base and equip people with the tools to operate within an environment shaped by machine learning.

RECOMMENDATIONS

Schools need to ensure that key concepts in machine learning are taught to those who will be users, developers, and citizens.

Government, mathematics and computing communities, businesses, and education professionals should help ensure that relevant insights into machine learning are built into the current education curriculum and associated enrichment activity in schools over the next five years, and that teachers are supported in delivering these activities.

In addition to the relevant areas of mathematics, computer science, and data literacy, the ethical and social implications of machine learning should be included within teaching activities in related fields, such as Personal, Social, and Health Education.

117. The Royal Society. 2014 Vision for science and mathematics education. See https://royalsociety.org/topics-policy/projects/vision/ (accessed 22 March 2017).

118. The British Academy. 2015 Count us in – Quantitative skills for a new generation. See http://www.britac.ac.uk/count-us-quantitative-skills-new-generation-bar (accessed 22 March 2017).

RECOMMENDATIONS

The next curriculum reform needs to consider the educational needs of young people through the lens of the implications of machine learning and associated technologies for the future of work.

An analysis of the future data science needs of students, industry and academia should be undertaken to inform future curriculum developments.

Informed users or practitioners of machine learning

Realising the economic benefits of machine learning will require a pipeline of people with high-level machine learning skills. In addition to supporting a broad understanding of the basic concepts in machine learning, more advanced understanding will be needed by a greater pool of people, to create informed users of machine learning.

Many professional occupations have been previously assumed to be relatively immune to the influence of new technologies, such as machine learning. However, as discussed in the following chapter, the growing capabilities of machine learning mean that professionals in a range of fields – such as medicine, law, accountancy, and engineering – will increasingly be required to make use of this technology in their day-to-day activities. For example, in the legal sector, machine learning is able to automate some of the research tasks of junior lawyers[119]. A range of occupations will therefore require practitioners to understand how to work with machine learning systems, and identify new areas where machine learning may be helpful in their professional lives. For example, in many scientific disciplines there will be opportunities for individuals with a deeper understanding of machine learning to use this understanding to advance their areas of research. Expanding the data skills in which scientists are trained to include machine learning could help facilitate this.

119. The Law Society. 2016 The future of legal services. See http://www.lawsociety.org.uk/news/stories/future-of-legal-services/ (accessed 22 March 2017).

At a minimum, newly-qualified professionals should be data literate, and able to 'think algorithmically'. An understanding of the capabilities of, and technology behind, machine learning is also necessary. Professional bodies can take an active role in addressing the changes brought about by machine learning technologies, and there are already examples of professional bodies that are considering the possible effects of these technologies[120].

Beyond these traditional professions, demand for practitioner-level skills in machine learning is likely across a range of industry sectors, as companies of all sizes seek to make use of this technology. If the UK is to fully reap the benefits promised by machine learning across these sectors, businesses of all sizes will need to be able to access a ready pool of talented people with machine learning skills. The demand for talented individuals with skills in machine learning in both tech-focussed and non-tech businesses is already high, and is likely to grow.

Online training courses – such as Massive Open Online Courses (MOOCs) – are increasingly being used to develop digital skills, and machine learning is an area of strength for these courses. Some also allow participants to work on machine learning problems within a business context[121].

Machine learning is a rapidly progressing area, and one which will be in demand across a range of fields. In the short term, new mechanisms are needed to build capability in machine learning, increasing the pool of people who have advanced machine learning skills.

120. See, for example: The Law Society. 2016 The future of legal services. See http://www.lawsociety.org.uk/news/stories/future-of-legal-services/ (accessed 22 March 2017) and The Association of Chartered Certified Accountants. 2016 Professional accountants – the future. See http://www.accaglobal.com/content/dam/members-beta/docs/ea-patf-drivers-of-change-and-future-skills.pdf (accessed 22 March 2017).

121. The Royal Society. 2016 Data analytics: the skills need in STEM. Conference report. See https://royalsociety.org/~/media/events/2016/11/data-science-workshop/data-analytics-conference-report-16112016.pdf (accessed 22 March 2017).

RECOMMENDATIONS

To equip students with the skills to work with machine learning systems across professional disciplines, universities will need to ensure that course provision reflects the skills that will be needed by professionals in fields such as law, healthcare, and finance in the future. Some exposure to machine learning techniques will also be useful in many scientific activities. Professional bodies should work with universities to adjust course provision accordingly, and to ensure accreditation schemes take these future skills needs into account.

In the short term, the most effective mechanism to support a strong pipeline of practitioners in machine learning is likely to be government support for advanced courses – namely masters degrees – which those working across a range of sectors could use to pick up machine learning skills at a high level. Government should consider introducing a new funded programme of masters courses in machine learning, potentially in parallel with encouragement for approaches to training in machine learning via MOOCs, with the aim of increasing the pool of informed users of machine learning.

Retaining the UK's leading position in the development of machine learning requires increased support for building advanced skills.

Developing and supporting the next generation of research leaders

There is already high demand for people with advanced skills in machine learning. Specialists in the field are highly sought after in the global market, and can command salaries accordingly.

This creates a challenge for academic research in machine learning; there is a growing range of companies which – recognising the value of machine learning to their business – are voracious consumers of talented machine learning researchers, often offering very attractive packages.

While there is nothing inherently wrong with this, the dominance of a small group – and therefore potentially a limited range of interests – risks skewing the field in a particular direction. Multinational corporations have resources to invest in machine learning, which go beyond anything the UK university sector could support. The dominance of these interests in driving machine learning research, both by providing funding and by attracting researchers, risks skewing the field in one direction, creating a so-called 'research monoculture'. This could leave interesting avenues of research unpursued (for lack of relevance to the sectors attracting talent) and fail to address key research challenges in areas of social concern.

This competition can also make it difficult to continue to recruit outstanding researchers into university posts, at all levels, which in turn has consequences for the future of advanced training and academic research leadership in the field.

If the UK is to remain at the forefront of developing this field, then further action is required to help cultivate advanced skills in machine learning, to support both academic and industrial advances. An effective public sector research funding environment should support this, taking a role in driving machine learning research which complements that within industry.

In part, this requires that existing funding structures are able to effectively support this type of developing technology. Important new disciplines often start out crossing traditional disciplinary boundaries, and this is the case for machine learning. This can raise impediments in a number of ways: lack of knowledge by potential students in making career choices; difficulties in competing for university curriculum space with more traditional fields; researchers and teachers spread across different university departments; and the potential for the new area to fall between gaps in existing research funding streams and committees.

Machine learning has grown out of advanced statistics, data science, and artificial intelligence, and today includes elements of each. It also feeds into applications across sectoral and research domains, with potential applicability in fields from mathematical to social sciences, from discovery to applied research, and in many commercial sectors. As an interdisciplinary area, machine learning is not currently well-served by existing funding models. Engineering and Physical Sciences Research Council (EPSRC) funding streams make provision for research into artificial intelligence, human-computer interaction, and robotics, amongst other potentially-related fields, but there is no single stream that 'fits' machine learning, so it can feel like there is not a natural home for this field in current funding structures.

This interdisciplinary nature will be of particular importance as the social consequences of machine learning become clearer. Effectively cultivating talent in this field will therefore require support from research funders and universities, to both ensure that institutional arrangements nurture the development of the field, and to ensure that its researchers have a broader understanding of the place of machine learning in society.

Another important contributor to building skills at this level is ensuring adequate funding for students working for advanced degrees. Recognising the significance of these types of advanced digital skills in helping to improve the UK's productivity, in the spring 2017 Budget, the Government reiterated its commitment to creating a highly-skilled workforce. To help create this workforce, the Government intends to create 1,000 more PhD places, and provide further funding for new fellowships for early- and mid-career researchers in areas aligned to the Industrial Strategy[122].

122. UK Government. 2017 Budget. 8 March 2017. See https://www.gov.uk/government/topical-events/spring-budget-2017 (accessed 22 March 2017).

RECOMMENDATIONS

In considering the allocation of additional PhD places and new fellowships across subject areas, as announced in the Spring 2017 Budget, machine learning should be considered a priority area for investment.

Because of the substantial skills shortage in this area, near-term funding should be made available so that the capacity to train UK PhD students in machine learning is able to increase with the level of demand from candidates of a sufficiently high quality. This could be supported through allocation of the expected 1,000 extra PhD places, or may require additional resources.

Universities and funders should give urgent attention to mechanisms which will help recruit and retain outstanding research leaders in machine learning in the academic sector. This academic leadership is critical to inspiring and training the next generation of research leaders in machine learning.

Technical capital

One element of technical capital relates to the availability of hardware to support machine learning systems. As noted in Chapter 1, one of the factors contributing to the recent proliferation of successful machine learning applications has been the availability of more powerful computers to support these analytical capabilities. Continuing to advance computing hardware in a way that enables more powerful analyses is an area of active research, and ensuring researchers have access to hardware that can support powerful machine learning techniques is an aspect of supporting technical capital in the field.

Access to software tools is also important, and there have been a number of developments in this space[123].

The importance of diversity in machine learning

Achieving excellence in science and technology requires a diverse and inclusive scientific workforce, which attracts talented people from a wide range of backgrounds, to create an inclusive and innovative environment. Ensuring that machine learning attracts this broad pool of people, with a range of perspectives, is therefore key to ensuring the ongoing health of the field, and encouraging continued innovation in this space.

The composition of the machine learning population is also significant in shaping how the field will advance. Those working in the field define the issues that are investigated, and the significance attached to different areas of work[124]. If the field is to advance in a way that represents a broad range of interests, then it will need to draw from a broad pool of people, if it is to avoid developing a form of research and development 'myopia'[125]. Attracting a range of people to the field will also be essential in improving the overall strength of the UK's skills base in this area.

Beyond their expertise in machine learning, machine learning developers have a strong role in shaping how the public will interact with machine learning. These interactions are shaped by the designs of the interfaces or apps with which people interact, as well as by technical properties of the algorithms and their training data. These aspects, in turn, are shaped by the developer, who will need to ensure machine learning can work for a broad range of people. Systems trained to work for, or with, only a narrow range of users are likely to encounter difficulties when used by the broader population, or have other unintended consequences[126]. A broader range of training data for these applications can help ensure their accuracy for a wide range of users.

Achieving excellence in science and technology requires a diverse and inclusive scientific workforce.

123. For example, the release of TensorFlow, an open-source software library developed by Google.

124. Turkle S. 1986 Computational reticence. In technology and women's voices (ed. C Kramarae). New York, US: Pergamon Press.

125. Clark J. 2016 Artificial intelligence has a 'sea of dudes' problem. *Bloomberg Technology*. See https://www.bloomberg.com/news/articles/2016-06-23/artificial-intelligence-has-a-sea-of-dudes-problem (accessed 22 March 2017).

126. Some of the unintended consequences of narrowly designed systems are described in: Crawford K. 2016 Artificial Intelligence's White Guy Problem. *New York Times*. 25 June 2016. See https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=0 (accessed 22 March 2017).

Machine learning has significant potential across industry sectors.

Mathematics, statistics and computer science, each include relatively low proportions of women participants; machine learning reflects these relatively low levels of diversity. For example, at 2016's Neural Information Processing Systems conference, the largest of the major international conferences in machine learning, only around 1,000 of the almost 6,000 attendees were women.

While it is difficult to find simple statistics that accurately represent the composition of the machine learning community, indications of this composition can be taken from the composition of related research communities. Women comprise 22.9% of higher education staff in the field of mathematics, and 22.2% of staff in IT, systems sciences and computer software engineering[127]. This gender disparity can also be seen earlier in the field; 17.1% of computer science and 38% of mathematical science graduates are women[128]. The field itself has recognised this disparity, as shown by the activities of groups such as Women in Machine Learning[129].

While data is less available on other aspects of diversity across machine learning, there also appear to be disparities in representation with regards to socioeconomic background and elements of BME representation.

4.2 Machine learning and the Industrial Strategy

Encouraging machine learning in business

The success of many of the largest and most rapidly growing global corporations has relied centrally or continues to depend on machine learning. Sundar Pichai, CEO of Google, said in 2016: "Machine learning is a core, transformative, way by which we are rethinking how we're doing everything. We are thoughtfully applying it across all our products, be it search, ads, YouTube, or Play[130]." Ginni Rometty, CEO of IBM, has said that her organisation is building the future of the company on machine learning[131]. This view of how broadly machine learning will affect software creation and delivery is widely shared by CEOs in the US technology industry[132]. There is thus a substantial sector of the growing digital economy that directly involves machine learning.

As outlined in Chapter 2, there is also a vast range of potential benefits from further uptake of machine learning across other industry sectors, arising from its use in streamlining existing processes, through to its potential to improve, or in some cases transform, these sectors. The economic impact of the technology could therefore play a central role in helping to address the UK's productivity gap[133]. This project has found significant potential for machine learning across manufacturing, pharmaceuticals, legal, energy, cities, and transport sectors.

127. Equality Challenge Unit (ECU). 2015 Equality in HE statistical reports 2015: part 1 (staff). See http://www.ecu.ac.uk/wp-content/uploads/2015/11/Equality-in-HE-statistical-report-2015-part-1-staff.pdf (accessed 22 March 2017).
128. Equality Challenge Unit (ECU). 2015 Equality in HE statistical reports 2015: part 2 (students) See http://www.ecu.ac.uk/wp-content/uploads/2015/11/Equality-in-HE-statistical-report-2015-part-2-students.pdf (accessed 22 March 2017).
129. Women in Machine Learning (WIML). See http://wimlworkshop.org/ (accessed 22 March 2017).
130. Levy S. 2016 How Google is remaking itself as a machine learning first company. *Backchannel*. See https://backchannel.com/how-google-is-remaking-itself-as-a-machine-learning-first-company-ada63defcb70#.ih8neznbn (accessed 22 March 2017).
131. See, for example: Nusca A. 2016 IBM's CEO thinks every digital business will become a cognitive computing business. *Fortune*. See http://fortune.com/2016/06/01/ibm-ceo-ginni-rometty-code/ (accessed 22 March 2017).
132. Executive Office of the President. 2016 Preparing for the future of artificial intelligence. Washington, US: The White House. 12 October 2016. See https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (accessed 22 March 2017).
133. UK Government. 2017 Budget. 8 March 2017. See https://www.gov.uk/government/topical-events/spring-budget-2017 (accessed 22 March 2017).

Sectors need access to a pool of skilled people, from those working to create innovative machine learning systems, to those who work at the interface of business and machine learning, to those who interact with the end-product of machine learning, or who need to be able to understand its foundations, such that they can spot the opportunities it presents. The demand for talented individuals with skills in machine learning in both tech-focussed and non-tech businesses is already high and exceeds the current pool of skilled individuals. In addition to the role of the skills system in providing this, government support systems for business will also play a role in helping companies to navigate the use of machine learning. For example, the Knowledge Transfer Network may have a role in enabling collaborations or increasing understanding of where and how machine learning can be used, which may be an area of interest for its Special Interest Group in Robotics and AI.

Experience at the interface between the mathematical and computational sciences and application areas suggests that it is vital to have individuals with deep knowledge of the sector, who then also have an understanding of machine learning, as well as having access to machine learning experts. It seems analogous to the situation in which many companies would have in-house legal expertise: individuals with legal training who know and understand the business and can deal directly with many issues, but who are also well placed to act as the interface with specialist external legal expertise when it is required. As noted earlier, one route to achieving this in the short term would be through masters-level training for appropriate individuals within sectors. This needs to be coupled with increasing the pool of machine learning specialists.

Increasing access to data which is in a machine-friendly form could catalyse further innovations based on machine learning, and help bring about its benefits. Such data can be made more accessible to machine learning systems by ensuring it is curated in a machine-readable way, based on the increased use of open standards. For instance, in an environment where data is used to support the functioning of cities – through transport, energy and services – a data infrastructure is needed. Such new infrastructure would define the data that needed to be collected, provide mechanisms to make this data open and usable, and ensure that it is produced in a way that is interoperable between systems. For companies operating in such an environment, data would be both an asset and a route to revenue, and used or shared accordingly.

Addressing key challenge areas in any one of many sectors – manufacturing, pharmaceuticals, the legal sector, energy, cities, and transport – could help secure a positive economic impact via machine learning.

Supporting entrepreneurship in machine learning: creating an environment that supports the commercialisation of ideas

A significant proportion of start-up companies cite machine learning as a competitive advantage, in most cases this is highly applied to their specific problem rather than general or generalisable. Given affordable access to and scalability of the technology, investors are often excited by this approach, so this trend is likely to continue; a few years ago start-ups cited 'big data' as their edge in a similar way.

The approach has been proven to be highly successful in many cases; as noted above, some of the largest global corporations have machine learning at their core. Many of these were relatively recent start-ups. This appears likely to continue to be an area which will drive economic growth.

Creating an environment in which start-up companies can thrive depends on many factors, including access to data, capital, ease of spinning-out from academic institutions, access to talent, access to computing power, and access to markets. The expansion of successful technology start-ups into major businesses or 'unicorns' (often defined as having a billion dollar valuation) depends on these and other factors. While there are groups that have grown organically to provide peer support for start-ups[134], government also has a role.

134. For example, meet-up groups as detailed.

Perhaps the most critical issue for machine learning start-up companies is human capital and talent. That talent is free to work or migrate anywhere in the world, and extensive global demand considerably exceeds supply. Success of UK start-up companies based around machine learning thus depends on the appeal of the UK, generally, and specifically the ecosystem for start-up companies, to allow them to attract and retain the best and brightest employees. The other key factor is the rate of creation of this talent, reinforcing the need for programmes to attract and train the next generation of machine learning experts. The difficulties of recruiting machine learning specialists can be especially acute for start-up companies. New companies will not have the reputational pull of the large tech companies, nor will they be able to offer potential recruits comparable packages. When combined with the lack of security in an early-stage company, recruiting employees in this key area can be especially difficult.

The supply of talent in machine learning is critical to the future competitiveness of the UK in this space, and Government has a role to play in ensuring that the UK is an attractive environment for such talent.

RECOMMENDATION

As it considers its future approach to immigration policy, the UK must ensure that research and innovation systems continue to be able to access the skills they need. The UK's approach to immigration should support the UK's aim to be one of the best places in the world to research and innovate, and machine learning is an area of opportunity in support of this aim.

The UK has nurtured high-profile and successful start-ups, and further action to support the start-up community could yield further economic benefits.

There does not currently appear to be a shortage of risk capital to support early-stage start-up companies in this field. But this may change as market conditions and sentiment evolve.

While many typical university spin-out companies would be based around a specific technological discovery, this standard model may fit less well for machine learning spin-outs. There may not be any IP *per se* to be licensed or transferred into a machine learning spin-out but rather know-how on the part of the academic founders that is central to the new business. Not having to negotiate IP issues can make the spin-out process simpler. Some universities may also think differently about the appropriate split of founding equity in this context. As in any spin-out, ensuring the largest possible alignment of incentives for the key stakeholders in the growing company can make a big difference to its future success.

Both the development and the training of machine learning algorithms can be extremely computationally intensive. Access to, and funds for, extensive computing can thus provide a competitive advantage. For example, one of several significant reasons given by executives at DeepMind for their decision to join Google was access to unrivalled computing resources. (The others were access to resources and an increased ability to attract the most talented workers). For most machine learning start-up companies, this can be a substantial challenge.

One direct way in which governments can potentially help particular start-up companies, where appropriate and allowable, is through their procurement processes. Government contracts help early-stage companies in several ways: they provide a source of income; they give the company the direct experience of engaging with customers, which provides important feedback for their developing market offering; and they act as external recognition of the company's product. This method for assisting early-stage companies appears more widespread in some competitor countries than it is in the UK.

In addition to general support for this important source of future economic growth, there may be benefits in critically assessing the UK's potential across this sector in order to identify potential value stacks: specific segments of the ecosystem where targeted investment or strategic support offers a particular opportunity because of the UK's strength and current position in that segment.

Strategic consideration should also be given to the right long term approach to maximising value from entrepreneurial activity in this space. On the one hand, the recent acquisitions of DeepMind, VocalIQ, Swiftkey, and Magic Pony, by Google, Apple, Microsoft, and Twitter respectively, point to the success of UK start-ups in this sector. On the other, they reinforce the sense that the UK environment and investor expectations encourage the sale of technologies and technology companies before they have reached their full potential. Selling medium-stage companies to large multinational corporations may ultimately reduce their impact on the UK economy when compared to a strategy which facilitates their growth into major independent corporations.

There are related but somewhat different challenges for start-up companies in other areas who wish to benefit from machine learning expertise. In particular, it is likely to be even harder for these companies to recruit machine learning expertise than for specialist machine learning start-ups. It is unclear whether these needs could be met through retaining consultants in machine learning, but in any case availability of such consulting support is also extremely limited[135].

The Industrial Strategy

To meet the demand for machine learning across industry sectors, the UK will need to support an active machine learning sector that capitalises on the UK's strengths in this area, and its relative international competitive advantages. This will draw from coordinated government action to support machine learning at all levels, including marshalling government procurement practices in a way that treats machine learning as a priority area for investment, supports UK-based machine learning businesses, and recognises the significance of machine learning in government support for business.

The Government has already recognised the significance of robotics and AI in supporting economic growth, via its new Industrial Strategy Challenge Fund. This fund will provide targeted support for research and development in a number of priority technologies[136]. The opportunities at stake – and the UK's existing strength in this field – make machine learning and AI key technologies for UK businesses, both now and over the next 5 – 10 years, for which support for research and development should be provided. This potential extends beyond the use of robotics in extreme environments[137], though machine learning will also be a key underpinning technology for such applications.

135. Some initiatives in this area have arisen organically, for example the London AI meet up community. See http://www.london.ai/ (accessed 22 March 2017).

136. Prime Minister's Office. 2017 Press release: PM announces major research boost to make Britain the go-to place for innovators and investors. See https://www.gov.uk/government/news/pm-announces-a-2-billion-investment-in-research-and-development (accessed 22 March 2017).

137. As specified by the Industrial Strategy Challenge Fund. See: UK Government. 2017 Budget. 8 March 2017. See https://www.gov.uk/government/topical-events/spring-budget-2017 (accessed 22 March 2017).

RECOMMENDATIONS

Government's proposal that robotics and AI could be an area for early attention by the Industrial Strategy Challenge Fund is welcome. Machine learning should be considered a key technology in this field, and one which holds significant promise for a range of industry sectors.

UK Research and Innovation (UKRI) should ensure machine learning is noted as a key technology in the Robotics and AI Challenge area.

In determining the shape and nature of Defense Advanced Research Projects Agency-style (DARPA-style) challenge funding for research, Government should have regard to facilitating the spread and uptake of machine learning across sectors.

Key sectors of UK industry – as outlined in this report – have the potential to adopt machine learning and create value from its use. However, uptake across sectors is patchy, and many areas of UK industry are not yet making use of this technology. For example, in manufacturing, pharmaceuticals, the legal sector, energy, cities, and transport there are challenges suitable for intervention, and potential for machine learning to disrupt current activities. Increasing the absorptive capacity of these sectors through the Industrial Strategy Challenge Fund could help deliver the benefits of machine learning more quickly, and Government should design challenges in these areas to push forward the use of machine learning accordingly.

Government needs to provide mechanisms to support people seeking to make use of machine learning through public support for entrepreneurism, small business, and enterprise.

Businesses need to understand the value of data analytics as a key part of business infrastructure. Government support for business should be able to provide advice and guidance about how to make best use of data, and organisations such as Growth Hubs or the Knowledge Transfer Network should ensure their business advisers are sufficiently informed about the value of data as business infrastructure to be able to provide guidance for businesses about, for example, the value of machine learning.

The Department for Business, Energy and Industrial Strategy (BEIS) should review support networks for small businesses to ensure they are able to provide advice and guidance about how to make use of machine learning, or to effectively support businesses offering machine learning products. This includes public-sector procurement processes, and the effectiveness of support for businesses using machine learning should be considered as part of the Government's review of the Small Business Research Initiative.

Chapter five
Machine learning in society

Left
Machine learning presents opportunities and challenges for society, which need to be navigated.

Machine learning in society

Continued public confidence in the systems that deploy machine learning will be central to its ongoing success.

While offering potential for new businesses or areas of the UK economy to thrive, the disruptive nature of machine learning brings with it challenges for society. Its technological capabilities enable new uses of data, which challenge existing data governance systems. Its new applications raise questions about public confidence and acceptability. Machine learning could bring significant benefits across a range of sectors, but careful stewardship will be needed to ensure that these are brought into being in a way that advantages all in society.

5.1 Machine learning and the public

Machine learning is already deployed in a range of systems or situations which shape daily life, whether it be by detecting instances of credit card fraud, providing online retail recommendations, or supporting search engine functions. As a result of this increasing pervasiveness, many of us will interact with machine learning-based systems every day, without necessarily realising what a powerful technology this is.

Yet experience of other emerging science and technology issues shows that early adoption does not guarantee continued support by all, or most, of the public. Furthermore, for a technology with such wide-ranging potential as machine learning, there is a risk that undesirable use in one area could undermine confidence in its use in other areas.

Continued public confidence in the systems that deploy machine learning will be central to its ongoing success, and therefore to realising the benefits that it promises across sectors and applications.

Prior to this project, evidence about the public's views on machine learning was scarce. Research on related topics has shown that:

- People are content with robots – which may draw on machine learning to carry out autonomous functions – being used in situations that could be dangerous or difficult for humans. However, people look less favourably on robots in more personal or caring roles, largely due to the fear of losing human-to-human contact[138].
- Attitudes towards the use of data – which is central to the development of machine learning systems – vary. There is generally low awareness of the potential uses of large datasets, and the extent to which people are supportive, or unsupportive, of their data being used for a particular application depends on the purpose, extent of anonymisation, and the potential public benefit[139].

From the start of this project, the Royal Society has sought to engage with the public to find out what their existing views on machine learning are, and their attitudes towards its applications. Through an ongoing programme of public events, the Society has been providing space for the public to find out more about machine learning. Early in the project, and in conjunction with Ipsos MORI, the Society carried out a public dialogue exercise to find out more about the public's views of machine learning. The questions raised during this study offer insights into how to create the conditions for rapid and safe delivery of the potential benefits of machine learning, while managing any associated risks.

138. Ipsos MORI. 2014 Public attitudes to science. See https://www.ipsos-mori.com/researchpublications/researcharchive/3357/Public-Attitudes-to-Science-2014.aspx (accessed 22 March 2017).

139. Ipsos MORI. 2014 Public attitudes to science. See https://www.ipsos-mori.com/researchpublications/researcharchive/3357/Public-Attitudes-to-Science-2014.aspx (accessed 22 March 2017) and Ipsos MORI. 2016 Public dialogue on the ethics of data science in government. See https://www.ipsos-mori.com/Assets/Docs/Publications/data-science-ethics-in-government.pdf (accessed 22 March 2017).

This research started with a quantitative survey of 978 people, which quantified current awareness of, and views about, machine learning, for a representative sample of the UK public. The results of this set out a baseline level of understanding and initial reactions to the technology.

Quantitative survey data were complemented, and given depth, by a dialogue process, in which members of the public and the Society's machine learning Working Group were brought together to discuss the implications of this technology. A series of dialogue events in Birmingham, Huddersfield, London, and Oxford provided a space where people could find out about machine learning, ask questions, share opinions, and develop their views. Through a series of case studies, participants could see the practical applications of machine learning, and deliberate about how it could be used in the future[140].

Recognition

Awareness of the term machine learning is low: only 9% of those surveyed had heard the term 'machine learning', and only 3% felt that they knew a great deal or fair amount about it. However, awareness of its applications is higher: 76% of respondents had heard of computers that can recognise speech and answer questions – this was the most frequently-recognised application – and 89% had heard of at least one of the eight examples of machine learning used in the survey. Awareness also varies demographically: respondents who were male, under 65 years old, and more affluent were more likely to say they had heard of machine learning.

These findings were echoed by participants at dialogue sessions: discussion about their current experiences of machine learning showed that they tended to have experienced the same sub-set of applications, such as recommender systems used by online retailers and offers from loyalty cards. Very few were aware of the mechanics of the technology used in these systems.

The low levels of public awareness of machine learning demonstrated in the quantitative survey increase the significance of a dialogue process in creating a space for people to develop their thinking about machine learning, and for gaining more in-depth insights into how people feel about the use of this technology in a range of settings. This process allowed exploration of the concerns and opportunities people saw associated with machine learning, and space to begin to develop a framework for how to evaluate this technology, as outlined in the sections which follow.

Attitudes to machine learning

One of the clearest messages from these public dialogues is that the public do not have a single view of machine learning; attitudes, positive or negative, vary depending on the circumstances in which machine learning is being used. As the application area provided the key lens through which the public evaluate uses of machine learning, general conclusions about the concerns or opportunities they saw associated with this technology are hard to define. There are themes that arise from discussions about different case studies, but the nature and extent of concerns, and the perception of potential opportunities, varied with application.

Although only 9% of people have heard of the term 'machine learning', the vast majority are aware of some of its applications.

140. These case studies considered the use of machine learning in healthcare, social care, marketing, transport, finance, crime and policing, education, and art.

Future work by the Royal Society

Building on the work of this project, including the *Hacking happiness* event[142] the Royal Society will be developing a series of activities to explore how AI could be used for new forms of social good.

Opportunities

The significant potential of machine learning was clear to many, not least because of its connection to the world of 'big data' and its ability to analyse data. By analysing this data, participants thought that machine learning could:

- be more objective than human users, or help avoid cases of human error, for example avoiding issues that may arise where decision-makers are tired or emotional;
- be more accurate, for example in detecting features in medical images and making accurate diagnoses;
- be more efficient, particularly in terms of public sector resources and shaping how services were delivered;
- offer opportunities for new businesses, and economic growth across a range of sectors; and
- play a role in addressing large-scale societal challenges, such as climate change or the pressures of an aging population[141].

People could therefore see machine learning improving how services work, saving time, and offering meaningful choice in an environment of 'information overload'.

Concerns

Concerns about machine learning and its applications fell into four themes:

- the potential for machine learning systems to cause harm, for example as a result of accidents in autonomous vehicles;
- the possibility that people could be replaced by machines in the workplace, or could become over-reliant on machines, for example in making diagnoses;
- the extent to which systems using machine learning might make experiences less personal, or human, either by changing the nature of valued activities, or by making generalised predictions about an individual; or
- the idea that machine learning systems could restrict the choices open to an individual, for example directing consumers to one type of product or service.

Each of these themes is currently being addressed by areas of active research – such as validation and verification, or human-machine interaction – progress in which could help increase public confidence in the deployment of machine learning systems (see also Chapter 6).

The key factors in each of these areas of concern are outlined in the subsections that follow.

141. Chapter 2 illustrates some potential applications of machine learning with broad societal benefits. Securing such benefits for society is also the focus of groups such as the Partnership on AI to Benefit People and Society.

142. A hackathon event held in partnership with the Digital Catapult, to explore how machine learning could be used in applications that sought to increase their user's happiness.

Harm

Central to many concerns about the use of machine learning was the fear that individuals would be harmed in the process, either directly – physical harm as a result of interacting with an embodied system – or as a result of the implications of a machine learning-driven system, such as misclassification or misdiagnosis.

The type of harm, results of this harm, and hence the strength of concern in this category, varied across different machine learning applications. There was some evidence to suggest that embodiment, and the creation of agents that act autonomously in the physical world, played a role in determining the extent to which harm was a key area of concern. This was most apparent in discussions about driverless vehicles or social care, where a physical agent operating independently tended to be associated with a greater risk of harm occurring.

In discussing what would give more confidence in the deployment of such systems, and what might therefore address concerns about potential harms, participants sought:

- reassurance that systems would be robust, with appropriate validation and testing;
- strong evidence of safety, and in some cases evidence that machine learning would be more accurate than humans carrying out an equivalent function; and
- in cases where the outcomes at stake were significant, some level of human involvement, either by making a decision on the basis of a machine's recommendation, or by taking an oversight role.

Replacement

Concerns that machine learning systems could replace humans were manifest in two ways.

Firstly, the potential impact of machine learning on employment was a clear area of concern, despite not being actively introduced into discussions of different use cases by facilitators. That this concern was raised spontaneously, and frequently, illustrates that it is an issue of high salience for the public. Participants in dialogue sessions could see clear links to previous advances in technology and the resulting impact of these on the workforce, for example citing the automation of car production lines as an example of how advancing technology had replaced human roles in the workforce.

And if previous technological advances had replaced elements of the workforce[143], the huge range of potential applications – a key opportunity – that people saw for machine learning intensified concerns about it also displacing human roles. Where previous advances in automation had affected a specific group, such as those involved in car production, one fear expressed was that the versatility of machine learning could cause mass unemployment.

Secondly, the applicability of machine learning to everyday activities prompted participants to question whether it would replace individual skills, such as reading a map or driving a car, which could be useful in everyday life. Such an over-reliance on technology, and potential de-skilling, raised questions about the ability of people to maintain effective judgement in situations where the relevant technology was not available.

People see clear benefits and opportunities associated with machine learning, but there are also concerns.

143. For example, from 1500 to 1800, the percentage of the British labour force working in agriculture fell from 75% to 35%. See: The Economist. 2014 The onrushing wave. See http://www.economist.com/news/briefing/21594264-previous-technological-innovation-has-always-delivered-more-long-run-employment-not-less (accessed 22 March 2017).

The extent and nature of these concerns are different in different contexts.

Impersonal experiences or services

A related concern was that of depersonalisation. While seeing potential for machine learning to be used to deliver a range of services or be applied in a range of situations, dialogue participants were concerned about what might be lost in the process.

For some, this was an intense reaction to the possibility of machine learning changing their relationship with an activity of personal significance; feelings of freedom or autonomy arising from driving a car, for example, or enjoyment taken from reading poetry, and relating to the person who wrote it. This reaction was closely linked to the development of specific applications in an area they considered to be integral to expressing their individuality or to their personal fulfilment. The prospect of machine learning – or perhaps any – technology taking over these particular activities was therefore unappealing[144].

For others, concerns about depersonalisation were connected to the delivery of key services, or to interactions with key personnel, frequently in caring roles or other scenarios where the ability to give an accurate response was not the sole measure of success. Qualities such as human empathy or personal engagement were generally desirable, and particularly important in areas such as health or social care. The prospect of reducing meaningful human-to-human interaction was therefore a concern. Robots might be able to perform certain tasks in social care, but this was acceptable as a means of freeing human carers to spend more time on other activities with the person for whom they were caring.

Restrictions on human experience

Participants understood machine learning as a technology that allowed analysis of vast quantities of data – more than humans could deal with – and use of this data to make decisions or predictions. However, some questioned the ability of machine learning to generate a nuanced interpretation: they believed machine learning would make broad generalisations, rather than individual predictions. Two areas of concern arose from this lack of confidence in the accuracy of machine learning systems, namely:

- that people could be mis-labelled, or inadvertently stereotyped, and have their activities mistakenly restricted as a result; and
- that machine learning could generate an algorithmic bubble, in which unusual or challenging opinions, experiences or interactions were filtered-out, ultimately narrowing the horizons of its users.

Again, the domain in which machine learning would be used was central to the seriousness of these concerns; the potential impact of being erroneously classified as high-risk for insurance products has obviously different consequences to receiving poorly-tailored shopping recommendations. Where the impact of mis-labelling has a significant influence on individual freedoms, finances, or safety, these considerations were more salient.

144. Conversely, the Government Office for Science's Foresight project on the Future of Identities makes the point that some forms of data-enabled technology allow people the option to create a sense of identity with others that they never had before, for example people with rare diseases. See: Government Office for Science. 2013 Future of identity. See https://www.gov.uk/government/publications/future-identities-changing-identities-in-the-uk (accessed 22 March 2017).

Public concerns are context specific

The application-specific nature of potential benefits and risks was key throughout our dialogue process; attitudes to the technology of machine learning itself were largely neutral. In evaluating the desirability of machine learning in different applications, participants took a broadly pragmatic approach, assessing the technology on the basis of:

- the perceived intention of those using the technology;
- who the beneficiaries would be;
- how necessary it was to use machine learning, rather than other approaches;
- whether there were activities that felt clearly inappropriate; and
- whether a human is involved in decision-making.

Accuracy and the consequences of errors were also key considerations.

Fundamentally, the concerns raised in these public dialogues related less to whether machine learning technology should be implemented, but how best to exploit it for the public good. Such judgements were made more easily in terms of specific applications, than in terms of broad, abstract principles.

RECOMMENDATIONS

Continued engagement between machine learning researchers and the public is needed: those working in machine learning should be aware of public attitudes to the technology they are advancing, and large-scale programmes in this area should include funding for public engagement activities by researchers. Government could further support this through its public engagement framework.

To help ensure those working in machine learning are given strong grounding in the broader societal implications of their work, postgraduate students in machine learning should pursue relevant training in ethics as part of their studies.

Future work by the Royal Society

The Royal Society has already carried out the first public dialogue exercise on machine learning. It will be building on this in coming years by creating spaces for further dialogue and interaction, and will be extending this engagement activity by working with partners including the museums sector and the media.

Machine learning is a new lens through which to view concepts such as consent and privacy.

5.2 Social issues associated with machine learning applications

The previous section described concerns raised in the Society's public dialogue exercise. This section considers some of these in more detail.

As it enhances our analytical capabilities, machine learning challenges our understanding of key concepts such as privacy and consent, shines new light on risks such as statistical stereotyping and raises novel issues around interpretability, verification, and robustness. Some of these arise from the enhanced analytical capabilities provided by machine learning, while others arise from its ability to take actions without recourse to human agency, or from technological issues. While machine learning generates new challenges in these areas, technological advances in machine learning algorithms also offer potential solutions in many cases. Chapter 6 proposes an agenda for machine learning research that would encourage the development of just such technological solutions related to each of the issues described below.

Use of data, privacy, and consent

Machine learning can be a tool to make sense of a new data environment, in which people and machines are continually networked, data is collected in new ways, and data use and re-use is increasingly dynamic. As it is put to use in this new environment, machine learning reframes existing questions about privacy, the use of data, and the applicability of governance systems designed in an environment of information scarcity.

The extent to which people are concerned about privacy and data use is variable both between individuals and for each individual between applications. For example, about 60% of people could be categorised as so-called 'data pragmatists', whose level of concern about data privacy depends on the circumstances at hand[145].

The analytical capabilities of machine learning in a 'big data' environment enable new uses of data, with datasets combined in innovative ways to draw new insights; fundamentally, its role is in creating information from data. Its capabilities therefore create a tension with privacy considerations, which may seek to hold back certain types of information.

In such an environment, the power of anonymisation techniques to preserve privacy is diminished. For example, by merging overlapping records which are in the public domain, advanced analytics can draw personal insights from open data. In one case, the personal health information of a US politician was discerned from a seemingly anonymised public database, through analysing this data in conjunction with other open health records and the voter registry[146].

Machine learning further destabilises the current distinction between 'sensitive' or 'personal' and 'non-sensitive' data: it allows datasets which at first seem innocuous to be employed in ways that allow the sensitive to be inferred from the mundane.

145. Ipsos MORI. 2017 Public views of machine learning: findings from public research and engagement (conducted on behalf of the Royal Society).

146. Ohm P. 2009 Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review* **57**, 1701–1777.

For example, research has shown how accessible digital records, such as Facebook 'Likes' by which Facebook users express positive sentiment about content on the social media site, can be used to infer sensitive personal attributes. By analysing a user's Facebook Likes, researchers found that they could predict characteristics such as sexual orientation, ethnicity, religious or political views, intelligence, or gender. Although Likes are publicly available by default, users do not necessarily expect these to reveal more sensitive information. Yet such information may now be predicted from their online activity, or digital records that are available to a range of organisations. Through seemingly innocuous online activity, people can therefore reveal more personal information, which they might be less comfortable sharing with corporations, governments, or other users of these websites[147].

There are already examples of how insights drawn from the use of advanced analytics on social media data might be used to make suggestions or decisions about the products or services offered to an individual. For example, in November 2016 the insurance company Admiral announced that it intended to analyse the Facebook posts of its customers – specifically aimed at 17 – 21 year olds – to gain insights into their anticipated driving behaviour, which it believed could be inferred from the content of their posts or Likes, and thereby create an idea of the level of risk they presented. Ultimately, this proposal was blocked by Facebook[148]. However, it demonstrates the ways in which data can be used, and reused, in novel ways.

In the past, consent has been pitched as the hallmark of good data governance. However, it is by no means clear that, even in cases where consent is used as the 'gold standard' for data use, this consent is informed. Although up to 33% of people claim they usually read website terms and conditions, server-side surveys indicate that only 1% actually have[149]. A consent-based approach to data governance relies on people having the understanding, time, and energy to invest in consenting. An increasingly dynamic data environment, in which data is re-purposed for previously unforeseen use further undermines the ability of individuals to meaningfully consent to their data being used. New approaches to navigating questions about consent are therefore needed.

The ability to draw connections between data is now so advanced that traditional approaches to managing privacy, such as de-identification, may no longer apply. Meanwhile, the balance of risks and benefits to the citizen as a result of these new uses of data may play out differently in different contexts such as healthcare or retail, muddying the waters with regards to what constitutes acceptable or unacceptable data use.

Questions about consent are further complicated by how ownership of different data types is perceived. Discussions in terms of 'my data' may lead to a data governance model based on specific consent, but does not reflect that much of the value of data comes from its combination, or, in the case of machine learning, in its use training an algorithm.

Challenges such as interpretability and accountability arise in some machine learning applications.

147. Kosinski M, Stillwell D, Graepel T. 2013 Private traits and attributes are predictable from digital records of human behaviours. *PNAS* **110**, 5802–5805.

148. Peachley K. 2016 Facebook blocks Admiral's car insurance discount plan. *BBC News*. 2 November 2016. See http://www.bbc.co.uk/news/business-37847647 (accessed 22 March 2017).

149. Ipsos MORI. 2014 Understanding society: the power and perils of data. See https://www.ipsos-mori.com/DownloadPublication/1687_sri-understanding-society-july-2014.pdf (accessed 22 March 2017).

Fairness and statistical stereotyping

Statistical profiling is already used in marketing, insurance, and assessment of threats for policing, so the need to carefully manage biases in data is not in itself new. There are two different ways in which machine learning applications may give rise to biases or lack of fairness.

The first occurs when machine learning algorithms inherit subjective biases which are present in the data on which the algorithms are trained. For example, if machine learning algorithms are used to screen job applications, they will typically need a large set of examples of job applications that have already been classified by humans into different categories (such as 'reject', or 'shortlist'). The machine learning algorithm will then look for features shared amongst the shortlisted applications, which help to discriminate them from the unsuccessful applications. If the humans making the initial decisions are themselves biased (even unconsciously), for example so that applications from men tend to be more successful than those from women, then the algorithm is likely to learn that an applicant's gender is one factor correlated with success, and incorporate that in its decision-making process. The resulting algorithm will be biased against women because it will have inherited this subjective bias from that of the human decision makers who classified the training data[150].

Biases arising from social structures can be embedded in datasets at the point of collection, meaning that data can reflect these biases in society. These effects can be quite subtle. They can arise from aspects of the way the data is collected (so-called ascertainment bias), or as consequences of conscious or unconscious biases in human decision-making which are reflected in training data.

A different source of bias or unfairness can arise when a machine learning algorithm correctly finds that a particular attribute of individuals is valuable in predicting outcomes, in contexts where society may deem use of such an attribute inappropriate. For example an algorithm which aims to predict mortgage default may find in its training data that, other things being equal, older individuals have an increased likelihood to default, and hence use an applicant's age in making lending recommendations. Although the association with age is real, society may decide that use of age information to decline mortgages is a form of age discrimination, and hence that it is not appropriate.

Where there are particular attributes that should not be used in decisions, it may not be enough simply to instruct the algorithm to ignore those attributes. Even if age, or race, or gender, are explicitly excluded from data used to make predictions, there may be good surrogates for these in the data. For example, information such as address, occupation, years in education, parents' birth places, may be quite predictive about race, so an algorithm which uses these indirect predictors can make different decisions on the basis of the race of the individual, even if ignores the direct information in the data about race.

150. See, for example: Miller C. 2015 Algorithms and bias: Q&A with Cynthia Dwork. *New York Times*. 10 August 2015 and Zemel R, Wu Y, Swersky K, Pitassi T, Dwork C. 2013 Learning fair representations. *JMLR Working and Conference Proceedings* **28**, 325–333.

Interpretability and transparency

Once trained, many machine learning systems are 'black boxes' whose methods are accurate, but difficult to interpret. Although such systems can produce statistically reliable results, the end-user will not necessarily be able to explain how these results have been generated or what particular features of a case have been important in reaching a final decision.

Where decisions or predictions have a significant impact – personally or socially – demonstrably higher accuracy than alternative techniques may not be enough to generate confidence in a machine learning system. In such contexts, understanding how the solution or decision was reached becomes more significant. From a technical perspective, increasing the interpretability of machine learning systems may also be desirable for several reasons.

First, interpretability and transparency can help people extrapolate an algorithm's behaviour to situations in which it has not been explicitly tested, thereby increasing confidence in its ability to perform well in a broad range of scenarios. While humans often have a good feel for how other humans will think and behave across a wide range of circumstances, we are understandably cautious about trusting an artificial algorithm which might employ very different types of analysis and may lack 'judgement brakes' that are implicit in human decisions. Transparency can also help in detecting instances of bias or unfairness.

Second, increased transparency – that is knowing when and why a system performs well or badly – may be directly helpful in the development of better algorithms. This can apply at the stage of tuning algorithms to improve performance during their development for particular applications. It can also help in understanding potential weaknesses of an algorithm. For example, a model designed for use in hospitals to predict the probability of complications or death as a result of pneumonia was found to be assigning pneumonia patients who also had asthma to a lower risk category than clinicians would have expected. Such patients were at higher risk of complications, but they also had higher survival rates: the model did not initially recognise that their seemingly lower risk stemmed from the greater medical attention and more intensive treatment that these patients received. If the model had been deployed without being able to examine how its inputs contributed to a decision, less aggressive forms of treatment would have been recommended for those patients, with potentially detrimental results[151].

Third, there may be situations in which society deems that principles of fairness require that an individual be given reasons when an important decision is made against them[152]. Where such a decision is made by an algorithm, this would require at least some level of transparency and interpretability of the algorithm. A 'right to an explanation' is implied in legal frameworks surrounding the use of data, namely the new European General Data Protection Regulation (see Box 4)[153]. When and whether such transparency should be required, exactly

151. Caruana R, You Y, Gehrke J, Koch P, Sturm M, Elhadad N. 2015 Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1721–1730.

152. See, for example: O'Neill C. 2016 Weapons of math destruction. London, UK: Allen Lane.

153. European Parliament and the Council of the European Union. 2016 EU General Data Protection Regulation – Recital 71. *Official Journal of the European Union* **59**, L119/1–L119/149.

what constitutes an appropriate explanation, and whether algorithms should be subjected to a different standard from human decision makers, are all complex questions with which society will need to grapple.

Since machine learning algorithms are encapsulated in computer programs, there is a sense in which publishing the program could be defined as explaining what the algorithm will do. Application of the algorithm will usually occur only after it has interacted with training data. Exactly because the algorithm learns from the training data, simply knowing the underlying program is different from knowing which features of the data the algorithm puts weight on in general, or in a particular instance. In our view, issues of transparency and interpretability cannot be resolved simply by making computer code available.

Approaches to model interpretability typically examine the algorithm, or consider the outcome, and may involve different methods of model validation. For example:

- Machines could be certified, demonstrating confidence that they achieve a certain level of competence.
- More advanced algorithms could create 'input-output' mappings of data, showing how different inputs influenced the output.

In attempting to resolve issues of transparency, there can be trade-offs between accuracy and interpretability. At a basic level, hard-coded rules are more interpretable, but more opaque approaches such as neural networks are often more powerful and can produce more accurate results. This trade-off between transparency and performance has different consequences in different applications, raising questions about whether the decision to prioritise transparency or accuracy needs to be made explicitly and, if so, how and by whom.

Such discussions also need to be framed in terms of the alternative; human decision-making can itself be somewhat opaque, prone to bias, or subject to a range of – very human – limitations, and it can also be based on expertise, skills, and experience.

Given the strengths and weaknesses of both machine-based and human decision-making systems, the extent to which a machine learning system constitutes an improvement on existing methods – reducing the number of road deaths, for example – may be key, and this will be highly context-specific[154].

154. There may be cases – for example, where gaming the system is an undesirable outcome – where a level of opacity is a virtue. See: Zarsky T. 2013 Transparent predictions. *U. Illinois Law Rev.* **2013**, 1503–1570.

BOX 4

The GDPR and a 'right to an explanation'

The General Data Protection Regulation – expected to take effect in April 2018 – is a set of EU regulations that, replacing the current Data Protection Directive, will set the framework for the processing of personal information. The Regulation does not apply unless the data at hand is both personal and being processed, though both of these terms are defined broadly.

Article 22 of the Regulation addresses "automated individual decision making, including profiling". It notes that the person to whom data relates "shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her" and that, if such processing is necessary, the person will have "at least the right to obtain human intervention [...] to express his or her point of view and to contest the decision[155]". Furthermore, Articles 13 and 14 of the Regulation require that, if profiling is used, the person to whom that relates should be able to access "meaningful information about the logic involved[156]".

Responsibility and accountability

There are related questions as to the contexts in which it is, or is not, appropriate for a decision to be made entirely by an algorithm with no human involvement. Results from public dialogue exercises indicated a general preference to see a 'human in the loop', that is human involvement in the final decision, for decisions with major impacts, although this may change as new systems become better established. On the other hand, although some people found directed internet advertising annoying, there was no suggestion that decisions about which advertisements to show, or which films to recommend, required human involvement. As for most issues, people's views depended greatly on the context and application area.

Automated decision-making systems are already in use, so in some senses these questions are not new. For example, an individual applying for a personal loan online might have their application assessed by algorithms and automatic searching techniques, in order to give an immediate yes or no response. The use of such systems is already governed by the Data Protection Act. This allows an individual to access information about the reasoning behind any decisions taken by an automated process, if there has been no non-automated involvement and the decision has had a significant impact.

155. European Parliament and the Council of the European Union. 2016 EU General Data Protection Regulation – Article 22. *Official Journal of the European Union* **59**, L119/1–L119/149.

156. European Parliament and the Council of the European Union. 2016 EU General Data Protection Regulation. *Official Journal of the European Union* **59**, L119/1–L119/149.

Future work by the Royal Society

Further work is needed to explore the ways in which machine learning might influence constitutional values, such as accountability. The Royal Society will be helping to shape discussions in this area through its high-level Science and the Law Programme, which will bring together senior scientists and members of the judiciary to explore the challenges associated with machine learning, liability, and accountability.

Who should be accountable when machine learning goes wrong, whether or not there is a human in the loop, and what recourse should individuals or groups have in this context? Machine learning offers the possibility of extending automated decision-making processes, allowing a greater range and depth of decision-making without human input. If this is to be achieved, and if it is not possible to explain why a machine learning system responded to its environment in a particular way, or how a situation arose, new models of accountability – and liability – may be required, or new ways of approaching the relationship between humans and software or machines. This becomes more complex in systems where chains of machine learning algorithms operate in tandem.

Amongst the public, the most common response to the question "who should be held accountable when machine learning 'goes wrong'" was "the organisation the operator and machine work for" (32%), and, though this did not constitute a majority response, it did clearly outweigh the number of respondents who believed the machine itself should be held accountable (3%).

This requirement for a narrative to help assign responsibility is, once again, application-specific: it may not matter that neither Alpha Go nor its human opponent could give a detailed explanation of the reasoning for their moves, but it will matter that there are ways to hold organisations in the public and private sector to account for a range of decisions.

It is already the case that decisions are made about people based on data, and some safeguards are in place around these. Questions about the accountability structures around automated decision-making are therefore not new. These questions continue to be germane in the context of machine learning, which also tests the boundaries of current approaches. Autonomous vehicles, for example, have prompted discussion about approaches to liability.

Society has yet to test the boundaries of current models of liability or insurance when it comes to new autonomous intelligent systems. Different approaches to addressing this issue have been suggested, and include (but are not limited to):

- The so-called Bolam Test, or whether a reasonable human professional would have acted in the same way;
- Strict liability – or liability without negligence or intent to harm – for autonomous vehicles; and
- Third party liability, akin to provisions made for dangerous dogs.

As case law in these areas develops, new models of liability may emerge, or gaps in existing legislative provisions may become clearer. In some application areas, this may also require new forms of insurance.

Potential social consequences associated with the increased use of machine learning

The increasingly pervasive use of machine learning raises a number of broader questions about its potential social consequences. Some of the live social issues provoked by increased use of machine learning are outlined below.

'Bubbles' and personalisation

Increasing personalisation, especially via social media tools which provide material – including news – that users are predicted to like and therefore click on, means that individuals can be consistently presented with a narrow version of the world, which is consistent with their current perspectives[157]. That this personalisation happens in different ways to different subsets of the population may: increase the likelihood of polarisation of perspectives; reduce an individual's awareness of the extent to which others might have a (very) different perspective; and decrease our ability to see or understand issues from the perspective of those in other 'bubbles'. There may be wider effects associated with changes to the way in which we interact with the world as a result of this personalisation, though it is not easy to predict what these are. One direct consequence may be a reduction in the extent to which we share experiences with many others in society[158].

New power asymmetries

The increased use of machine learning raises issues about how society is structured in a world where algorithms are widely used. In doing so, new power asymmetries may be created; a 'Faustian pact' whereby individuals willingly give up privacy in exchange for efficiency, convenience, or a need to access a service, without giving informed consent. This may be a result of the inaccessibility of terms and conditions, or because the consequences of consenting are difficult to predict.

Privacy could be thought of as not only an individual right, but a collective good. In a world where this is widely given up, there are likely to be societal consequences, though it is hard to foresee exactly what these will be[159]. Given the nature of the market in which technology companies operate, the relationship between citizens and companies in this space needs to be considered. The Royal Society's and British Academy's project on the governance of data and its uses will be considering the role of different areas of legislation in creating an effective data governance framework.

Human-machine interaction

As these new capabilities for computer systems become increasingly mundane, our relationship with – and expectations of – the technologies at hand will evolve. This raises questions about the long-term effects upon – or expectations from – people who have grown up with machine learning systems and smart algorithms in near-ubiquitous usage from an early age.

While some of these issues are likely to have social consequences, the impact of machine learning on society will likely be subtle, and it is at this point not possible to predict with any certainty what those consequences would be.

There are questions about the potential social consequences of increasingly pervasive machine learning systems.

157. Recent debates about the political narratives and perspectives to which people are exposed on social media – especially during elections – may be one example of this.

158. Turkle S. 2013 Alone Together. New York, US: Basic Books.

159. Yeung K. 2017 Algorithmic regulation: a Critical Interrogation (talk at the Sackler Forum on the Frontiers of Machine Learning). See http://www.nasonline.org/programs/sackler-forum/frontiers-machine-learning.html (accessed 22 March 2017).

5.3 The implications of machine learning for governance of data use

Ensuring the best possible environment for the safe and rapid deployment of machine learning will be essential for enhancing the UK's economic growth, wellbeing, and security, and for unlocking the value of 'big data'. Such deployment will depend in part on continued public confidence in machine learning technologies.

In considering issues related to governance of machine learning it is important to emphasise from the outset that by governance we include the whole configuration of legal, ethical, professional, and behavioural norms of conduct, conventions and practices that, taken together, govern the collection, storage, use, and transfer of data. It is also helpful to distinguish two separate, but related, spheres. The first is in relation to the way in which machine learning algorithms make use of, and interact with, data sets upon which they are trained. As already noted, the power and extent of machine learning challenge many current notions around privacy, consent and appropriate data use. The second sphere relates to questions around the properties of the resulting algorithms, after they have been trained on data, including safety, reliability, interpretability, and responsibility.

Although not specifically driven by machine learning, the volumes, portability, nature, and uses of data in a digital world raise many challenges for which existing data access frameworks do not seem well equipped. It would appear timely to consider how best to address these novel questions via a new framework for data governance. This is an area in which there is already substantial work being undertaken by the Royal Society and British Academy as part of a separate report due in summer 2017.

In the context specifically of machine learning, data governance may and should address questions as to whether a specific data set or type of data can be used for a particular purpose, and by whom. It may also consider whether different data sets or different types of data can be combined, at all or for a particular purpose. It should consider which safeguards are needed to minimise risks to individuals. Questions about the details of the algorithm used seem separate from those of data governance. For example, if a data governance body deemed that it was appropriate for a particular data set to be used by a specific organisation to make decisions about mortgage lending, it could and should confirm that the proposed use abides by the relevant standards, but subject to this it should not matter whether the algorithm employed for this task uses decision trees or neural nets (two different machine learning methods).

Once the machine learning algorithm has been trained on data it will then be applied. As noted earlier, in some uses this training of the algorithm is undertaken once, whilst in others the training is updated after each use. There are important social questions in some contexts about the standards that should be required of the trained algorithm, such as whether it needs to be interpretable (and if so what is meant by this) or whether it should only be used when there is a human in the loop. Society's answers to these questions may well differ across different settings. Such questions also seem separate from those around whether particular data sets can be used for certain purposes.

In contrast to this need for a new framework for data governance, there are many reasons why it does not seem appropriate to consider a new governance framework for machine learning *per se*.

First, machine learning algorithms are just computer programs, and the range and extent of their use is extremely broad and extremely diverse. It would be odd, unwieldy, and intrusive to suggest governance for all uses of computer programming, and the same general argument would apply to all uses of machine learning.

Second, in many or most contexts machine learning is generally uncontroversial, and does not need a new governance framework. How a company uses machine learning to improve its energy usage or warehouse facilities, how an individual uses machine learning to plan their travel, or how a retailer uses machine learning to recommend additional products to consumers would not seem to require changes to governance. (It should of course be subject to the law, and also involve appropriate data use, an issue to which we return below.)

Third, many of the issues around machine learning algorithms are very context specific, so that it would be unhelpful to create a general governance framework or governance body for all machine learning applications. Issues around safety and proper testing in transport applications are likely to be better handled by existing bodies in that sector; questions about validation of medical applications of machine learning by existing medical regulatory bodies; those around application of machine learning in personal finance by financial regulators.

We have seen that there are significant and important unresolved issues for some applications of machine learning, such as whether algorithms need to be interpretable in particular use cases, when humans should be involved in decision processes, and when algorithms should be held to a higher standard of accuracy or interpretability than human decision-makers. The answers to these questions will vary with the application area. This application-specificity is key when considering machine learning: some applications may require regulation to ensure public confidence, while others will be non-controversial. Some may be dealt with adequately via existing mechanisms. Others may need new frameworks, but these should be context specific, rather than being provided by an overarching governance system for machine learning.

There are also existing laws that govern the use of data and algorithmic decision-making. These existing regulatory mechanisms may apply to some existing applications of machine learning. However, these were written in an era when many of the applications of machine learning had not been conceived. The field is also changing rapidly, with an increasing number of applications and changing technological capabilities

There are technological responses to some of these challenges, as outlined in the following chapter, which will influence the nature of these trade-offs in future. Further technical advances could therefore create their own solutions, instead of relying on other regulatory mechanisms.

RECOMMENDATIONS

There are governance issues surrounding the use of data, including those concerning the sources of data, and the purposes for which it is used. For this, a new framework for data governance – one that can keep pace with the challenge of data governance in the 21st century – is necessary to address the novel questions arising in the new digital environment. The form and function of such a framework is the basis of a study by the Royal Society and British Academy.

It is not appropriate to set up governance structures for machine learning *per se*. While there may be specific questions about the use of machine learning in specific circumstances, these should be handled in a sector-specific way, rather than via an overarching framework for all uses of machine learning; some sectors may have existing regulatory mechanisms that can manage, while in others there may not be these existing systems.

5.4 Machine learning and the future of work

Machine learning and automation

Machine learning is enabling the automation of an increasing range of functions, which until recently could only be carried out by humans. While debates about the impact of technology – and automation in particular – are not new, the nature of these debates are now changing, as the capabilities of machine learning develop and it supports automation of a broad range of tasks.

These capabilities now extend beyond routine work, with machines able to exercise what might be thought of as judgement or creativity. This means that, unlike previous waves of automation, these new advances can be applied to a broader range of functions that are currently carried out by humans. They expand the influence of machine learning to a range of occupations, including roles which were previously thought to be relatively immune to automation.

At present, machine learning is automating the routine technical tasks in many fields, but the applications of machine learning in these areas are diversifying, from machine learning-powered chatbots giving free legal advice, to medical apps using machine learning to monitor health. Whilst alleviating the burden of some mundane tasks, this could affect employment and progression within a wider range of fields, which could require new approaches to staff training and development.

By performing certain tasks to a higher standard than humans, machine learning has the potential to improve productivity, increase efficiency, and ensure consistency of service. In some cases, this type of automation relies on virtual agents or systems, while others also draw from robotics in developing new physical systems. Machine learning is already automating a range of tasks:

- Routine administration: the London Borough of Enfield is using a so-called 'cognitive agent[160]' – or virtual employee – to carry out routine administrative tasks, such as handling requests for permits, authenticating licenses, or responding to routine queries from residents. This virtual agent, called Amelia, uses machine learning to analyse the content of these queries, and respond accordingly.
- Responding to customer queries: Hilton Hotels is developing an automated concierge service – Connie – which can respond to guests' questions[161].
- Responding to customer queries: Ocado uses machine learning to help organise responses to customer queries. Its natural language processing system scans the contents of emails, categorising and prioritising these on the basis of the type and urgency of query[162].
- Writing news stories: the Associated Press is using machine learning systems to automatically generate news stories about company performance. Based on analysis of corporate reports, and using style and content preferences, machine learning is used to extract key facts from corporate reports, and use these to create a news story[163].
- Promoting online news: machine learning systems can predict how popular or widely shared topics, or similar news content, will be on social media. For example, the New York Times created a bot called Blossom, which predicts how successful articles will be, and suggests which stories should be promoted[164].
- Financial trading: traders can use machine learning algorithms to scan news stories, financial information and press releases, and use the facts or sentiments contained therein to make predictions about the future performance of companies, in order to inform trading decisions[165]. Increasingly sophisticated automated trading algorithms are making use of machine learning to automate a range of decisions.

Machine learning enables automation of a range of tasks currently carried out by humans.

160. Davies W. 2016 Robot Amelia – a glimpse of the future for local government. *The Guardian*. 4 July 2016. See https://www.theguardian.com/public-leaders-network/2016/jul/04/robot-amelia-future-local-government-enfield-council (accessed 22 March 2017).
161. IBM. 2016 Hilton and IBM Pilot "Connie", the world's first Watson-enabled hotel concierge. See https://www-03.ibm.com/press/us/en/pressrelease/49307.wss (accessed 22 March 2017).
162. Voice A. 2016 How Ocado uses machine learning to improve customer service. 13 October 2016. See http://www.ocadotechnology.com/our-blog/articles/How-Ocado-uses-machine-learning-to-improve-customer-service (accessed 22 March 2017).
163. Madigan White W. 2015 Automated earnings stories multiply. *The Associated Press*. See https://blog.ap.org/announcements/automated-earnings-stories-multiply (accessed 22 March 2017).
164. Wang S. 2015 The New York Times built a Slack bot to help decide which stories to post to social media. *NiemanLab*. See http://www.niemanlab.org/2015/08/the-new-york-times-built-a-slack-bot-to-help-decide-which-stories-to-post-to-social-media/ (accessed 22 March 2017).
165. Knight W. 2016 Will AI-powered hedge funds outsmart the market? *MIT Technology Review*. See https://www.technologyreview.com/s/600695/will-ai-powered-hedge-funds-outsmart-the-market/ (accessed 22 March 2017).

Machine learning can augment the roles humans carry out.

- Giving legal advice: the DoNotPay chatbot is a machine learning system which helps its users to challenge the validity of parking tickets. The bot guides users through the relevant areas of the law, using machine learning to ask questions that can help determine whether the ticket can be appealed, for example on the basis of obscured signs or extenuating circumstances[166].
- Carrying out legal research: machine learning can be used to interrogate large legal databases as part of the early-stage legal research process. In these systems, natural language processing is used to review the contents of documents and databases, detecting patterns and creating outputs that can be interrogated further by lawyers[167].
- Reviewing medical images: machine learning systems can be used to analyse tissue sample images and detect features of disease. Such systems work by detecting features of anomalies in images, and can be used to improve the accuracy of diagnostics[168].
- Translating text or speech into another language: machine learning can process text and translate this into another language. This function is being put to use in smart phone apps, but also used to automatically translate documents into different languages for international organisations.
- Automating warehouses: machine learning can be used to support advances in robotics that automate some of the key tasks in warehouse management, notably packing goods and moving these around the warehouse.
- Driving: machine learning is a key technology supporting driverless vehicles. It can be used to help these vehicles navigate road systems, by recognising obstacles or road signs, or by adapting driving style on the basis of environmental conditions. Such vehicles are being developed by many companies, and include autonomous buses and lorries.

The examples above illustrate the broad impact of machine learning systems, which could result in changes to the world of work. However, while there are some tasks that machine learning systems can carry out more accurately or effectively than humans, there are many others where human competencies remain higher, and which may not ever be adequately replicated by machines. In the short- and medium-term, machine learning algorithms will be able to carry out specific tasks with accuracy and efficiency at levels comparable to or better than humans. In many settings it is likely that these will impact on specific aspects of employment, in augmenting what humans do, rather than in replacing entire roles.

166. The world's first robot lawyer. See http://www.donotpay.co.uk/ (accessed 22 March 2017).

167. The Law Society. 2016 The future of legal services. See http://www.lawsociety.org.uk/news/stories/future-of-legal-services/ (accessed 22 March 2017).

168. Gulshan V *et al*. 2016 Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. **312**, 2402–2410. (doi: 10.1001/jama.2016.17216)

The impact of machine learning on employment

Much has been written about the potential impact of machine learning, AI, and automation, on the economy, and on employment. Widely quoted figures include:

- 35% of jobs in the UK could have more than a 66% chance of being automated over coming decades[169].
- Up to 15 million jobs in the UK could be automated over the coming decades[170].
- It is technically possible to automate over 70% of the component tasks for 10% of jobs in the UK today[171].
- Up to 30% of jobs in the UK may be susceptible to automation by the 2030s[172].

And yet no single study has been able to capture the nuances of how machine learning will pervade the world of work in the coming decades, or when these changes might happen. In considering these estimates, analysts have variously noted the need to consider jobs that might be created, how changes might affect different sectors differently, the new ways in which people and machines will work together instead of substituting for each other, or whether the myriad of proposed applications of machine learning will be economically feasible to roll out in the near term.

For example, one estimate suggested the age of big data created 58,000 new jobs per annum from 2012 to 2017[173]. Another suggested that – while displacing over 800,000 jobs in this period – technology created over 3.5 million new jobs from 2001 to 2015[174]. Meanwhile the prediction that 35% of UK jobs were at risk of automation further found that this risk played out differently across different sectors: wholesale and retail had the greatest overall numbers of job at risk of automation, with 59% of current jobs having a high chance of being automated in the next two decades (2,168,000 jobs), with the figure for transport and storage being 74% (1,524,000 jobs), and health and social work 28% (1,351,000 jobs)[175]. Furthermore, the type of job within these sectors also influences its likelihood of automation: one estimate suggests that jobs carried out by workers educated to secondary school level were 15 times more likely to be automated than those carried out by workers with PhDs or masters degrees[176]. However, there is significant variability across roles, with factors

Common ground on the nature, scale, and timing of potential changes to the world of work as a result of machine learning is hard to find.

169. Frey C, Osborne M. 2013 The future of employment: how susceptible are jobs to computerisation? *Technol. Forecast. Soc.* **114**, 254–280.
170. Haldane A. 2015 Labour's Share (speech given to the Trades Union Congress). 12 November 2015. See http://www.bankofengland.co.uk/publications/Pages/speeches/2015/864.aspx (accessed 12 March 2017).
171. Arntz M, Gregory T, Zierahn U. 2016 The risk of automation for jobs in OECD countries: a comparative analysis. *OECD Social, Employment and Migration Working Papers* **189**, 34. (doi:10.1787/5jlz9h56dvq7-en)
172. PwC. 2017 UK economic outlook: Consumer spending prospects and the impact of automation on jobs. See http://www.pwc.co.uk/services/economics-policy/insights/uk-economic-outlook.html (accessed 22 March 2017).
173. The British Academy. 2015 Count us in – Quantitative skills for a new generation. See http://www.britac.ac.uk/count-us-quantitative-skills-new-generation-bar (accessed 22 March 2017).
174. Deloitte. 2016 Press release: Automation transforming UK industries. 22 January 2015. See https://www2.deloitte.com/uk/en/pages/press-releases/articles/automation-and-industries-analysis.html (accessed 22 March 2017).
175. Deloitte. 2016 Press release: Automation transforming UK industries. 22 January 2015. See https://www2.deloitte.com/uk/en/pages/press-releases/articles/automation-and-industries-analysis.html (accessed 22 March 2017).
176. Arntz M, Gregory T, Zierahn U. 2016 The risk of automation for jobs in OECD countries: a comparative analysis. *OECD Social, Employment and Migration Working Papers* **189**, 34. (doi:10.1787/5jlz9h56dvq7-en)

such as management, expertise, interfacing with people, and performing physical activities in unpredictable environments each reducing the likelihood of activities being automated[177].

Common ground on the nature, scope, and scale of the impact of AI on employment is difficult to establish: different AI technologies can be put to use in different ways to automate different tasks in different fields and to different timelines, in addition to creating new types of work or opportunities for human-machine collaboration. Given the difficulty of this task, it is unsurprising that conclusions from existing studies into the impact of these technologies on employment differ so substantially.

Through the varying estimates of jobs lost or created, tasks automated, or productivity increased, there remains a clear message: machine learning will have a significant impact on the way we work, and its effects will be felt across the economy. Machine learning is likely to become increasingly pervasive, and will affect everyone as it does so.

What is less clear, however, is whether the changes that arise as a result of the impact of machine learning will be like-for-like, with new tasks created, or whether certain tasks, roles, or people will be displaced.

Machine learning will increasingly feature in both our work and personal lives. While not necessarily replacing jobs or functions outright, machine learning will force us to think about our occupations, and the skills necessary to function in a world where these systems are ubiquitous.

The impact of machine intelligence on work is an issue that is already permeating the public consciousness. Replacement by machines emerged as a key concern from our public dialogue exercises with Ipsos MORI, despite participants not having been prompted directly on this point. In this context, the broad range of potential applications of machine learning contributes to both its benefits and its risks; participants questioned whether machine learning could potentially drive replacement of workers on a large scale – and across sectors – in a way that affected both skilled and manual workers. In contrast to technological changes in the past, which affected specific sectors, people see machine learning as driving more sweeping changes in how labour is organised. In tandem, participants were also concerned that increasing 'intelligence' could foster over-reliance on technology, with people de-skilling in certain areas – for example, if medical professionals were relying on computers for diagnoses – as a result.

177. Manyika J, Chui M, Miremadi M, Bughin J, Geogre K, Willmott P, Dewhurtst M. 2017 Harnessing automation for a future that works. McKinsey Global Institute. See http://www.mckinsey.com/~/media/McKinsey/Global%20Themes/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works-Full-report.ashx (accessed 22 March 2017).

These concerns are also felt by those working in the field of AI, robotics, and economics: one survey on the potential impact of automation on employment to 2025 found that only 52% of experts in these fields predicted optimistic future scenarios; 48% expressed concern[178].

Machine learning has the potential to disrupt the way in which value is created, and how the economic benefits of this value are distributed. Society is now in the foothills of the broader changes that AI will bring to the world of work over the coming decades.

It is highly likely that it is not just machine learning, but machine learning alongside other data-based techniques and advances, such as those in robotics, that will be disruptive[179]. In attempting to understand the current landscape and interpret how decisions about machine learning made today might affect its future, taking account of the growing body of insight into how emerging technologies are viewed and used as they move from novel to mainstream may be helpful[180]. The nature, scale, and duration of this disruption will depend on the social, political, ethical, and legal environments in which these technologies evolve.

Who benefits from AI-driven changes to the world of work will be influenced by the policies, structures, and institutions in place. Understanding who will be most affected, how the benefits are likely to be distributed, and where the opportunities for growth lie will be key to designing the most effective interventions to ensure that the benefits of this technology are broadly shared. To avoid creating a group of people who are left behind by the advance of this technology, action is needed to develop policy responses that will enable citizens to adapt to this new world of work.

At this stage, it will be important to take 'no regrets' steps, which allow policy responses to adapt as new implications emerge, and which offer benefits in a range of future scenarios. One example of such a measure would be in building a skills-base that is prepared to make use of new technologies, through increased data and statistical literacy, as discussed in Chapter 3.

178. Smith A, Anderson J. 2014 AI, robotics, and the future of jobs. Pew Research Centre. See http://www.pewinternet.org/2014/08/06/future-of-jobs/ (accessed 22 March 2017).

179. See, for example: Piketty T. 2013 Capital in the twenty-first century. Boston, US: Harvard University Press, on the shift from wealth based on physical assets to wealth based on labour, access to data, and skills to exploit it.

180. See, for example, Edgerton D. 2006 The shock of the old: technology and global history since 1900. Oxford, UK: Oxford University Press.

It is important to think now about how the benefits of machine learning can be shared.

Future work by the Royal Society

There will be an enduring question about how machine learning and AI change the way we all work. The Royal Society will continue to explore the potential impact of machine learning on work.

Evidence from our dialogue exercise demonstrates that fears about job losses arising from AI and automation are already present in the public consciousness. However, in addition to these concerns, the public could see how machine learning systems could improve the world of work and articulated a desirable relationship between people and machine learning systems.

The productivity dividend

While the exact nature and extent of the impact of machine learning on employment is not currently clear, it is clear that this technology will change the world of work, and will affect a broad range of jobs. It is also clear that the increased adoption of machine learning methods will be driven by the increases in productivity which they provide.

Previous major waves of technological change, including the industrial revolution, the use of electricity, and the development of electronics, have also been characterised by productivity increases. In each case, the benefits of these productivity increases have been spread. There have been benefits across society through raised living standards and wellbeing, as well as substantial financial benefits to a small subset of individuals or corporations. There have also been changes in the work environment with some jobs or sectors being lost, or substantially changed, and others being created.

In the same way, there will be a 'productivity dividend' generated by machine learning, in parallel with changes to the world of work and other aspects of people's lives. What is not clear is how the productivity dividend will be shared and who the major beneficiaries will be. At one extreme, much of the benefit may go to a small number of individuals or companies, with others losing jobs or facing reduced living standards; some inequalities will increase in this scenario in ways consistent with what is often attributed to 'globalisation'. At the other extreme, active steering through social choices would direct the productivity dividend more evenly across society, and perhaps specifically to those most adversely affected by increased use of machine learning, for example in specific employment sectors.

At this early stage of the cycle it may be possible for society to shape the way the productivity dividend is shared, for example by engaging industry in discussions about how the demand for new skills can be met (and funded), and other policy responses to ensure there are not groups of people left behind as a result of social changes to which this technology contributes. The potential benefits accruing from machine learning and their possibly significant consequences for employment need active management. Without such stewardship, there is a risk that the benefits of machine learning may accrue to a small number of people, with others left behind, or otherwise disadvantaged by changes to society.

While it is not yet clear how potential changes to the world of work might look, active consideration is needed now about how society can ensure that the increased use of machine learning is not accompanied by increased inequality and increased disaffection amongst certain groups. Thinking about how the benefits of machine learning can be shared by all is a key challenge for all of society.

RECOMMENDATION

Society needs to give urgent consideration to the ways in which the benefits from machine learning can be shared across society.

Chapter six
A new wave of machine learning research

Left
In addition to areas of research which address technical challenges, an exciting new wave of machine learning research is developing, which can advance the field's technological capabilities, while helping to address societal challenges. © cybrain.

A new wave of machine learning research

A new wave of machine learning research can both push forward its technical capabilities, while addressing areas of societal interest.

6.1 Machine learning in society: key scientific and technical challenges

Machine learning is a vibrant field of research, with a range of areas for further development across different methods and application areas. Recent advances in the field have opened up further exciting research challenges. Some challenges are of a largely technical nature, including, for example:

- Creating algorithms that can computationally scale to large data sets;
- Designing algorithms that do not require large amounts of labelled data;
- Designing data-, energy- and other resource-efficient machine learning methods;
- Advancing generative models that can be put to use in simulations; and
- Improving hardware to support powerful machine learning systems.

In addition to these purely technical challenges, the increasing range of applications of machine learning has opened up a new wave of research challenges, which relate to both technical advances in the field and societal challenges associated with machine learning. If addressed, these areas of research could accelerate the development of the field, while working to ensure continued public confidence in machine learning systems.

This chapter highlights a set of research topics where progress would have a direct impact on areas of public and societal concern around machine learning. Progress in these areas is essential if machine learning applications are to fulfil their promise in ways which continue to protect important principles and values in our society. The list which follows is not intended to be exhaustive.

6.2 Interpretability and transparency

Can we create machine learning systems whose workings, or outputs, can be understood or interrogated by human users, so that a human-friendly explanation of a result can be produced?

Increasing the interpretability of machine learning methods is desirable for a number of reasons, as noted earlier. These include the need to understand the processes used in safety critical systems or the ways in which decisions about individuals have been reached.

There are different possible approaches to achieving interpretability.

Machine learning methods could be restricted to those that directly yield an interpretation which is easy for humans to understand. One example of such an approach is a decision tree, which repeatedly makes sequential decisions according to simple rules. However, a significant drawback to this approach is that there may be important trade-offs between interpretability and accuracy. Further, if only repeated simple decision rules are allowed, then in order to make accurate predictions, it may be necessary to apply many thousands of rules, thereby losing the desired feature of interpretability.

A more nuanced approach involves tackling a classification or prediction task as a pipeline of machine learning models. The output from each model is fed as input into the subsequent model, such that the detection of generic features in the original input and formulation of a classification or prediction based on these features is split into two or more stages. The benefit of this approach is that the intermediate outputs of models in the pipeline can be designed so as to be interpretable by humans.

This gives additional scope to analyse the sequence that leads to specific outputs in response to given inputs. However, this approach will not be applicable in all cases.

Another approach is to proceed in two stages. First, a prediction system is designed solely to optimise accuracy. Then, a second system is trained specifically to explain the predictions and operation of the first system. There is promising recent work in this direction. One idea is to explore predictions locally; another is to examine the sensitivity of predictions to input variables. However, by its nature, this approach is essentially approximating an explanation for the first model, and hence may fail to explain behaviour or lead to false confidence in some important settings.

A more elaborate approach would be to create an interface between machine learning systems and human-machine dialogue systems so that, in the future, humans could talk to the machine and interrogate its reasoning. This may seem very appealing, but it will rely on underlying explanatory ability combined with confidence in the speech interface, where ambiguities might creep in and lead to potential misunderstanding.

BOX 5

Why are some machine learning systems 'black boxes'?

A neural network is an approach to machine learning in which small computational units are connected in a way that is inspired by connections in the brain. These systems may consist of many layers of neurons: the base layer receives an input from an external source, then each layer beyond it detects patterns in activity from the neurons in the layer beneath, integrates these inputs, and then passes a signal to the next layer[181]. In this way, signals can be passed through many layers, before reaching a top layer where a decision about the input is made. So, if the initial input is an image, the initial signals might come from the pixels of the image, and the top-level decision might be what object is in the image. As such systems learn from data, they strengthen or weaken synaptic connections to make their outputs more accurate.

This approach to processing data means that information across the neural network is highly dispersed, with complicated patterns of connection strength between units, and with potentially many thousands of layers of units. The result is the so-called 'black box' issue: these systems can create highly accurate results, but it is difficult to explain why a result has been obtained.

However, not all machine learning methods use this approach, and alternative approaches can be more readily interpreted.

181. Jones N. 2014 Computer science: the learning machines. *Nature* **505**, 146–148. (doi:10.1038/505146a)

6.3 Verification and robustness

Can we create more advanced, and more accurate, methods of verifying machine learning systems so that we can have more confidence in their deployment?

In many applications – especially safety-critical applications – the quality of the decisions or predictions made by machine learning systems needs to be verifiable to a high standard.

There are already ways of testing the robustness of machine learning systems. For example, so-called training / test splits of data can be used to verify predictions, by training an algorithm using a sub-set of the available data, and then putting it to use on another sub-set of the dataset to test its results (the 'test' data).

However, this methodology suffers from at least two limitations. Firstly, when deployed in the 'real world', the system may encounter data that falls far outside the range of that covered by the test data. This means that the conclusions drawn from testing might not be generalisable upon deployment. In the first instance, this is addressed in the formulation of training and test datasets, by aiming to ensure that they are drawn from as broad a distribution of available data as possible, with as close a match as possible to key features of real world data.

Secondly, in the case of online machine learning systems, where learning algorithms continue to be applied to the trained model after deployment, the system adapts in response to interaction with its environment. Both its behaviour, and the data to which the system is exposed, can change. This creates a complex pattern of interactions, which creates difficulties for formally guaranteeing the performance of the system.

Several lines of research are seeking to increase the robustness of machine learning systems in novel environments, and develop methods of verification.

One approach is to guarantee a minimum level of performance, by optimising the system to work on the theoretically worst possible observable data during the training phase. This can give some sense of security, and may even help combat adversarial attacks, but its overly cautious approach could lead to poor performance in typical conditions.

Another approach is to create systems that are context aware. A system which can monitor its own actions and its environment can compare these to the conditions under which it was trained; this allows the system to proceed with the method with which it was trained – if the context matches – or raise an alert and use more cautious approaches, if the context is outside of the ordinary.

Verification and robustness of machine learning systems is particularly important as a result of the potential impact of perturbations to these systems. When using machine learning, small changes to a system can be quickly replicated and deployed, with effects on a large-scale.

6.4 Privacy and sensitive data

What are the technical solutions that can maintain the privacy of datasets, while allowing them to be used in new ways by different users?

Privacy is an important theme in the deployment of machine learning systems, since such systems rely on large amounts of data to make inferences, predictions, and decisions. In many contexts, notably around data on individuals, it will be important to respect the privacy of the data. These issues relate to both the raw data used by machine learning systems, and to the derived or inferred data generated by these systems.

While some scenarios require absolute protection of privacy, good data governance in others may focus on the balance between a very small risk to privacy and the potential benefits to be gained from particular uses of the data. In both these settings, technological approaches themselves can be helpful in minimising risks to privacy.

One approach to privacy is de-identification of data, in which personally identifiable characteristics are removed from datasets. Sometimes this may substantially reduce the value of the data. In other examples, it may fail to protect privacy because identification of individuals can be possible using indirect cues, even when directly identifying data have been removed. Research is thus needed on more subtle and more sophisticated approaches to this question.

Advanced solutions to privacy questions include:

- Differential privacy, which randomly perturbs data, so that any individual's record is obscured, while amalgamated features are mostly left intact; and
- Homomorphic encryption, which allows computations to be performed on fully encrypted data, so that the raw data is never seen by the machine learning system.

Research in these areas is at an early stage, working with very simple calculations, and it is not yet clear if it will be possible to scale these approaches to useful levels of complexity.

Limits on access to data can restrict the development of corresponding machine learning systems. Often, valuable data lies in the hands of corporations or governments, which are unlikely to give access to researchers or practitioners for reasons that might include privacy or security concerns, legal constraints, reputational risk, or (for corporations) competitive advantage.

Technological solutions can play a role in increasing access to data in these complex situations. One possibility is through sharing only aggregated summaries of the data. Others could exploit privacy preserving machine learning systems, or distributed systems that run on each client's own personal device, sharing minimal information.

6.5 Dealing with real-world data: fairness and the full analytics pipeline

How can real world data be curated into machine-usable forms, addressing 'real-world' messiness, and potential systemic – or social – biases?

Machine learning systems need to contend with the realities of real-world data: data sets often have missing entries and outliers, they come in different formats, and suffer from various forms of data corruption.

Arguably most of the time spent on the practical deployment of machine learning systems is consumed by data cleaning, understanding, transformation, and integration. Yet this stage of the full data analytics pipeline receives relatively little research attention. Better methods to automate the full pipeline are needed, as well as clear audit trails of processes used to transform the data to ensure that these have not distorted results. Since each sub-component in a system introduces its own errors and inaccuracies, extensive processing can lead to unforeseen difficulties.

Further research in this area could develop more rigorous approaches to quantify the level of introduced errors, for example through:

- Developing standards for the processing and sharing of data, which allow the data quality to be assessed.
- Developing standards for evaluating the behaviour of systems and recognising the presence of undesirable biases and errors.

Technological solutions can also help ensure machine learning systems handle data fairly, and in ways that are in accordance with anti-discrimination legislation. For example, machine learning systems can be coded in a way that restricts how they use different inputs. As noted earlier, this does not, of itself, exclude possible discrimination because there may be indirect surrogates of the excluded inputs from which they can be predicted. A more restrictive approach would insist that machine learning algorithms produce outputs which are demonstrably uncorrelated with the characteristics related to anti-discrimination issues. For example, if an algorithm were just precluded from using the ethnicity of individuals, it may still be discriminatory because its conclusions may use other variables correlated with ethnicity, such as address, income, family size, and job. The stronger approach would restrict attention to algorithms whose conclusions were demonstrably uncorrelated with ethnicity. There is a need for research to understand how to develop algorithms with this property.

6.6 Causality

How can machine learning methods discover cause-effect relationships, in addition to correlations?

Machine learning systems extract statistical relationships from data, showing correlations between variables. However, cause-effect relationships are usually of more significant interest, both because they help to increase understanding of what the data means, and, critically, as they can be used to make decisions about interventions.

Discovering cause-effect relationships is a particular challenge, as there may be hidden biases in how data was collected. This has been a longstanding interest in cognate disciplines such as statistics. With the advent of sophisticated machine learning algorithms, it would be timely to focus research on whether and how these approaches can inform on potential causality.

6.7 Human-machine interaction

How do we design machine learning systems so that humans can work with them safely and effectively?

Many machine learning systems are and will be deployed in situations where they interact with humans, or in settings where the data with which they interact is not static. This presents a number of technical challenges, opportunities and concerns, for example:

- How do we best combine human intelligence and machine learning?
- How do we ensure that machine learning systems will perform as expected with humans in the loop?
- How do we design effective decision support tools based on machine learning?

Systems are already being developed to try to read the emotional state of a human, and respond accordingly in order to best address their needs. While this will provide many benefits, an important concern relates to how this may lead to undesirable influence of systems over humans. Even now, news sources and social media sites may tailor the stories and adverts that an individual is shown in order to maximise revenue. When these technologies are combined with perhaps more powerful, emotional channels, it is prudent to consider the effect on society. For example, one might imagine that data might show that angry people buy more of a certain type of product; this could lead a profit-maximising entity to promote anger-inducing stories. While such behaviour may already be part of our economic environment, machine learning-optimised channels of influence may raise the level of concern.

A related area of great interest is understanding how emotion helps humans to deal effectively with scenarios, and how these benefits might be incorporated into an artificial system. For example, through reinforcement learning, it is natural for an agent to weigh up the value of 'curiosity' when considering whether to invest additional resources into exploring potential actions and their consequences.

In addition to these areas of human-computer interaction, there are areas of research to explore how humans and AI systems can work together in partnership. To create effective partnerships, it is necessary to understand the strengths of each partner, and design systems with these in mind[182].

6.8 Security and control

How do we ensure machine learning systems are not vulnerable to cyber-attack?

As devices, such as phones or household appliances, becoming increasingly 'smart', with the ability to collect and analyse data and communicate with other devices or control units (the so-called 'Internet of Things'), they will be able to respond more intelligently to our needs.

However, in addition to the concerns above about data privacy, there are valid worries about the security of these devices, and what might happen if a malicious user (or 'virus' type algorithm) were to gain control of an increasingly large and inter-connected part of our environment.

The Royal Society's report *Progress and research in cybersecurity* noted how technical and social change required new approaches to cybersecurity, and that progressing these would require substantial research and development. Such research might include:

- New technologies and models for Internet of Things network security, which would reflect their highly distributed nature and the need for security by design.
- New resource-constrained crypto- and multi-factor authentication technologies for Internet of Things.
- Other cybersecurity challenges, as outlined in the Royal Society's report on cybersecurity[183].

182. Jennings N, Moreau L, Nicholson D, Ramchurn S, Roberts S, Rodden T. 2014 Human-agent collectives. *Commun. ACM.* **57**, 80–88.

183. The Royal Society. 2016 Progress and research in cybersecurity. See https://royalsociety.org/topics-policy/projects/cybersecurity-research/ (accessed 22 March 2017).

Further research is needed to safeguard our systems and ensure that they remain under control of the rightful user. Research into safety and control could seek to address the unintended consequences of machine learning systems, and the design and control of these systems to ensure this.

6.9 Supporting a new wave of machine learning research

Machine learning should be considered a priority area for investment in science, research, and innovation[184]. In this chapter, we have outlined a collection of specific research areas where progress would directly address areas of public concern around machine learning or constraints on its wider use. Research in these areas could be encouraged and supported through existing funding mechanisms. Alternatively, a series of challenges that would fund machine learning research in areas relevant to topics of social interest could advance the field of machine learning itself, and ensure this advancement was done in a way that increased confidence in machine learning systems.

RECOMMENDATION

Progress in some areas of machine learning research will impact directly on the social acceptability of machine learning in applications and hence on public confidence and trust. Funding bodies should encourage and support research applications in these areas, though not to the exclusion of other areas of machine learning research. These areas include algorithm interpretability, robustness, privacy, fairness, inference of causality, human-machine interactions, and security.

184. Such areas are discussed in: HM Government. 2017 Green paper: building our industrial strategy. See https://beisgovuk.citizenspace.com/strategy/industrial-strategy/supporting_documents/buildingourindustrialstrategygreenpaper.pdf (accessed 22 March 2017).

Annex / Glossary / Appendices

Left
Machine learning can be used in agriculture, for example in systems that identify weeds in crop fields using image recognition, and target them for removal.

Annex

Canonical problems in machine learning

The field of machine learning investigates the mathematical foundations and practical applications of systems that learn from examples, data, and experience. It draws from a range of disciplines, including computer science, statistics, engineering, and cognitive science.

Canonical problems in machine learning – the fundamental problems that machine learning seeks to solve – relate to: classification, regression, clustering, semi-supervised learning, and reinforcement learning.

This annex expands on Table 1, setting out some of the applications and techniques used to tackle these canonical problems.

1. Classification

Classification involves taking data and assigning it to one of several categories. The task at hand is to predict discrete class labels from input data, after a model has been trained on labelled data.

Applications of classification include face recognition, image recognition, and medical diagnosis. Typical methods for such tasks include: Logistic Regression, Support Vector Machines, Neural Networks, Random Forests, and Gaussian Process Classifiers.

2. Regression

Regression analyses try to predict continuous quantities from input data. Its applications include financial forecasting, and click rate prediction, which includes a range of applications in internet advertising.

Typical methods to address this task include Linear Regression, Neural Networks, and Gaussian Processes.

3. Clustering

Clustering is used for analysis where there is a lot of data, which needs to be organised in a way that creates clusters where similar points are grouped together.

This is used in bioinformatics and studies of gene expression, astronomy, document modelling, and network modelling. Typical methods include k-means, Gaussian mixtures, and Dirichlet process mixtures.

4. Dimensionality reduction

Dimensionality reduction is used in applications where the raw data has a high number of dimensions. These approaches map high-dimensional data onto low dimensions, while preserving relevant information, and have a range of applications, for example in data mining, scientific analysis, or image recognition. Typical methods include: Principal Components Analysis, Factor Analysis, Multidimensional Scaling, Isomap, and Gaussian Process Latent Variable Models.

5. Semi-supervised learning

In analyses where lots of unlabelled data is available alongside a few data points which have been labelled, for example a small number of annotated images, semi-supervised learning can be used to combine and learn from each of these.

This approach is particularly useful in applications where labelling data is expensive, for example in drugs trials, or other studies involving complex experiments. Methods include probabilistic models, graph-based semi-supervised learning, and transductive Support Vector Machines.

6. Reinforcement learning

Reinforcement learning addresses tasks where an agent or computer program needs to learn to interact with its environment, receiving inputs, and making sequential decisions so as to maximise future rewards. It therefore relates to adaptive control, and sequential decision-making under uncertainty.

Agents using reinforcement learning might be physical or virtual, and applications are found in robotics, games, trading, and dialogue systems. Methods in this field include Q-learning, direct-policy methods, and PILCO.

Glossary

Algorithm
A set of rules a computer follows to solve a problem.

Artificial intelligence
An umbrella term for the science of making machines smart.

Bayes' theorem
A theory that specifies how to handle uncertainty by updating the probability for a particular event, phenomenon, or hypothesis in response to data.

Bias (sampling)
Selection of data or samples in a way that does not represent the true parameters (or distribution) of the population. Bias in training data leads to bias in algorithms: machine learning is a data driven technology and the characteristics of the data are reflected in the properties of the algorithms.

Big data
Large and heterogeneous forms of data that have been collected without strict experimental design. Big data is becoming more common due to the proliferation of digital storage, the greater ease of acquisition of data (e.g. through mobile phones) and the higher degree of interconnection between our devices (i.e. the internet).

Data
Numbers, characters or images that designate an attribute of a phenomenon

Deep learning
A machine learning method which composes details together to obtain more abstract, higher level, features of the data through composition of mathematical functions. Powerful modern deep learning algorithms often involve a large number of these levels.

Differential privacy
An approach to protecting an individual's data by 'corrupting' it with noise before processing with the algorithm.

Gaussian
A probability density which adopts a 'bell curve' shape (and its generalisation to higher dimensions). It is widely deployed due to computational advantages and the tendency of independent data corruptions, when added, to be distributed according to this density.

Governance
The institutional configuration of legal, ethical, professional and behavioural norms of conduct, conventions and practices that, taken together, govern the collection, storage, use and transfer of data and the institutional mechanisms by and through which those norms are established and enforced.

Machine intelligence
A general term for machines that have been programmed to be smart, or otherwise artificially intelligent.

Machine learning
A set of rules that allows systems to learn directly from examples, data and experience.

Metadata
'Data about data', contains information about a dataset. For example, this information could include why and how the original data was generated, who created it and when. It may also be technical, describing the original data's structure, licensing terms, and the standards to which it conforms.

Model
A mathematical description of a system.

Neural network
A computer model with a particular form that was originally inspired by early work on understanding the nervous system.

Petabyte
1,000 terabytes or 10^{15} bytes of information

Program
A set of instructions given to a computer to allow it to carry out a task.

Reinforcement learning
An approach to machine learning in which an agent learns to interact with its environment, receiving inputs, and making sequential decisions so as to maximise future rewards. An important feature in this context is that it is often only after the agent makes a number of decisions that it learns of the payoff resulting from the set of choices. One challenge in reinforcement is thus to work out which of the decisions were "good" and which less so.

Sensitive (data)
Sensitivity has strict definitions under the Data Protection Act, but for the purposes of this report it refers to data or information that an individual would not wish to be widely and openly known or accessible.

Supervised learning
An approach to machine learning which relies on training data that has been labelled, often by a human. A label could be a categorisation into one or more groups: this is known as classification.

Test data
Data that is used to test the functioning of a machine learning system, or verify its outputs.

Training data
Data that can be used to train machine learning systems, having already been labelled or categorised into one or more groups.

Unsupervised learning
An approach to machine learning that uses data which has not been labelled. Commonly it will seek to determine characteristics that make the data points more or less similar to each other and will attempt to represent the data in a summary form, such as through clusters or common features.

Appendix

Working Group members

The members of the Working Group involved in this report are listed below. Members acted in an individual and not a representative capacity, and declared any potential conflicts of interest. Members contributed to the project on the basis of their own expertise and good judgement.

Chair	
Professor Peter Donnelly FMedSci FRS	Professor of Statistical Science and Director of the Wellcome Trust Centre for Human Genetics, University of Oxford
Members	
Professor Margaret Boden FBA OBE	Research Professor of Cognitive Science, University of Sussex
Professor Roger Brownsword	Professor of Law, King's College London
Professor Zoubin Ghahramani FRS	Professor of Information Engineering, University of Cambridge Chief Scientist, Uber (from March 2017)
Dr Nathan Griffiths	Associate Professor, University of Warwick
Dr Demis Hassabis	Founder and CEO, Google DeepMind
Dr Sabine Hauert	Assistant Professor, University of Bristol
Hermann Hauser KBE FREng FRS	Entrepreneur Co-Founder, Amadeus Capital Partners
Professor Nick Jennings FREng	Professor of AI, Imperial College London
Professor Neil Lawrence	Professor of Machine Learning, University of Sheffield Director of Machine Learning, Amazon (from September 2016)
Professor Sofia Olhede	Professor of Statistics, University College London
Professor Marcus du Sautoy FRS	Professor of Mathematics, University of Oxford
Professor Yee-Whye Teh	Professor of Statistical Machine Learning, University of Oxford Research Scientist, Google DeepMind (from August 2016)
Professor Dame Janet Thornton DBE FRS FMedSci	Director Emeritus, European Bioinformatics Institute

Royal Society staff

Staff from across the Royal Society contributed to the production of this report.

Royal Society staff	
Dr Claire Craig CBE	Director of Science Policy
Dr Natasha McCarthy	Head of Policy – Data
Jessica Montgomery	Senior Policy Adviser and Project Lead
Tracey Hughes	Head of Marketing and Public Engagement
Dr Franck Fourniol	Policy Adviser
Susannah Odell	Policy Adviser
Will Kay	Policy Intern
Previous Royal Society staff who contributed to the development of the project	
Tony McBride	Director of Science Policy (until December 2015)
Dr Nick Green	Head of Projects (until February 2016)
Belinda Gordon	Senior Policy Adviser (until November 2015)
Aleks Berditchevskaia, Amelia Dearman, Clare Dyer, Fiona McLaughlin	Policy Interns (various periods)

Review Panel

This report has been reviewed by an independent panel of experts, before being approved by the Council of the Royal Society. The Review Panel members were not asked to endorse the conclusions or recommendations of the report, but to act as independent referees of its technical content and presentation. Panel members acted in a personal and not a representative capacity, and were asked to declare any potential conflicts of interest. The Royal Society gratefully acknowledges the contribution of the reviewers.

Review Panel	
Mike Lynch OBE DL FRS FREng	Founder, Invoke Capital
Professor Genevra Richardson FBA	Professor of Law, King's College London
Professor Chris Williams	Professor of Machine Learning, University of Edinburgh
Professor Tom Simpson	Associate Professor of Philosophy and Public Policy, Blavatnik School of Government

Participants

The Royal Society would like to thank all those who contributed to the development of this project through submission of evidence and attendance at events.

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society. These priorities are:

- Promoting excellence in science
- Supporting international collaboration
- Demonstrating the importance of science to everyone

For further information
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2500
E science.policy@royalsociety.org
W royalsociety.org

Registered Charity No 207043

ISBN: 978-1-78252-259-1
Issued: April 2017 DES4702