MOTT MACDONALD

# Artificial intelligence

## Opportunity or threat?

July 2018

10 Fleet Place
London EC4M 7RB
United Kingdom

T +44 (0)20 7651 0300
F +44 (0)20 7248 2698
mottmac.com

# Artificial intelligence

## Opportunity or threat?

July 2018

# Issue and revision record

| Revision | Date | Originator | Checker | Approver | Description |
|---|---|---|---|---|---|
| 0.2 | 20/06/18 | A. Wheen | S. Clayton-Mitchell | | Initial draft for review |
| 0.3 | 25/06/18 | A. Wheen | | | Update following initial review |
| 0.4 | 17/7/18 | A. Wheen | R. Ramchurn | | Update following website review |
| | | | | | |
| | | | | | |
| | | | | | |

**Information class: standard**

# Contents

# 1 Introduction

You have probably noticed that advertisements appearing on the web have a knack of knowing exactly what you are interested in. You may have sought help from a digital assistant such as Alexa or Siri, and you will almost certainly have spoken to intelligent machines on the telephone. Whether you are aware of it or not, artificial intelligence (AI) is playing an increasingly important part in your life.

Much of the recent AI research and development has been driven by technology companies, but applications for AI can be found across a far wider range of businesses. Companies wishing to assess the potential of the new technology can make use of cloud-based AI platforms from service providers such as IBM, Microsoft, Google and Amazon. Many of Mott MacDonald's clients can already deploy AI to gain competitive advantages for their businesses.

This document provides an introduction to AI and describes a range of applications for the technology. It also considers the new threats that are being created by AI, and how these should be addressed.

## 1.1 Background

In 1951, Alan Turing proposed a test to determine whether a machine is as intelligent as a human. By 1958, some of the key components of artificial intelligence (AI) – such as speech recognition, language translation, image recognition and decision-making – had been identified. However, despite dramatic progress since then, no machine has yet passed the Turing Test.

Early attempts to build intelligent systems used a rules-based approach. Each problem was broken down into a series of yes/no decisions that could be implemented in software, but it soon became clear that it is simply impractical to develop software to answer every possible question, and the decision-making process implemented by the software must be able to evolve over time to reflect new insights and changing attitudes. While a rules-based approach works well for clearly-defined problems that have clearly-defined answers, it is much less effective for problems that require a degree of judgement.

As a result, AI research switched from a rules-based approach to a machine-learning approach in which the computer is 'trained' with a huge amount of data to show it the responses that are likely to be appropriate when faced with a particular problem. The computer has no 'understanding' of the decision-making process, but it can arrive at a suitable answer by correlating the question with other questions to which it does know the answer. To the computer, it's all about analysing probabilities.

Neural networks provide a way of implementing a machine that learns by imitating some of the functions of the human brain. They are constructed from layers of interconnected, neuron-like nodes that can carry out simple mathematical operations. Once the neural network has been 'trained' using an appropriate set of data, it can then be used to solve new problems.

This machine-learning approach means that a computer can sometimes deal with problems that it has not been trained to handle so long as the solution can be inferred from similar problems that it has encountered before. Humans have the same ability, but the computer cannot

normally produce a chain of reasoning to explain how it reached a particular solution. Consequently, artificial intelligence is very different from human intelligence, and the two forms of intelligence are often complimentary rather than competitive.

For some time now, it has been possible to teach AI systems to play games such as chess and Go by getting them to play against human experts. However, this approach is unlikely to enable the machines to move far beyond the capabilities of their human opponents, so Google's AlphaGo system uses a different approach; it teaches itself. In the space of three days, it advanced from a basic understanding of the rules of Go to the point where it was inventing sophisticated moves that were unknown to human experts. By May 2017, AlphaGo had beaten the world's best Go player and achieved a higher ranking than any human.

AlphaGo has demonstrated that it is not always necessary to transfer human expertise to an AI system if the problem that it is trying to address (eg: becoming a better Go player) can be clearly expressed by a set of rules. However, other AI applications require the machine to interact with humans, and an AI system cannot always distinguish between good and bad influences. Microsoft demonstrated this very clearly in 2016 when they launched a 'chatbot' on the internet to investigate how AI can engage with young web users in casual conversation. The system was designed to learn from the users interacting with it, but it very quickly started repeating the racist, sexist, homophobic and anti-Semitic attitudes that it acquired from those users. After 16 hours, the stream of invective had become so intolerable that Microsoft was forced to terminate the experiment. Although this is an extreme example, the datasets used to train an AI system have to be chosen carefully.

The success of AlphaGo illustrates that machines are now capable of inventing better machines without the need for human assistance. While machines have outclassed humans in terms of their physical capabilities for several thousand years, they are now also starting to move beyond humans in terms of their problem-solving capabilities. This will deliver some major benefits but it also creates some significant new threats. These will be discussed in the following sections.

# 2 Applications

The range of AI applications is already very wide and will continue to expand. A few examples are described below.

## 2.1 Interactive services

Internet technology giants such as Google, Facebook and Amazon collect large amounts of data about their users. They are able to do this because they make extensive use of automation driven by AI. Having a much better understanding of individual interests and preferences enables them to deliver highly-targeted advertising to each individual user, and this form of advertising is extremely valuable.

Many online services are delivered via web interfaces, but voice interactivity is becoming increasingly viable. During a developer conference in May 2018, Google demonstrated the Duplex AI helper which can make phone calls where, in certain circumstances, it convincingly passes for a human being. This has sparked a debate about whether 'chatbots' should be required to warn users that they are not human. At the moment, customer service transactions typically use AI in conjunction with a human, but it does not require much imagination to see that additional work is likely to migrate from humans to computers. Some direct marketing calls are already fully automated.

## 2.2 Medical

The pattern-matching capabilities of AI systems can be used to detect and diagnose illnesses. For example, a machine can check radiology scans for signs of disease and, unlike a human, it won't lose focus as a result of boredom or tiredness. Although machines already diagnose certain medical conditions more reliably than a human doctor, a human is typically required to check the machine's conclusions.

However, in April 2018, the US Food & Drug Administration approved an AI system for diagnosing diabetic retinopathy. This is the first time that the agency has approved a machine to provide a screening decision without the need for a doctor to confirm the results. The FDA has hinted that other AI-based diagnostic systems could be approved in the near future[1].

## 2.3 Predictive maintenance

Monitoring the moving parts in critical equipment such as trains, aircraft engines and large water pumps can enable mechanical failures to be predicted. Once the warning signs have been detected, intervention can be arranged to rectify the problem before the failure actually occurs.

Predictive models can also identify other problems that are likely to occur simultaneously or shortly after the original failure, thereby enabling the true impact of the fault to be assessed[2]. This means that AI systems can optimise the scheduling of engineering staff to minimise the cost of repairs while maintaining service quality and meeting safety requirements. In effect,

[1] "FDA approves AI-powered diagnostic that doesn't need a doctor's help", Emily Mullin, www.technologyreview.com, 11th April 2018.

[2] "Exposing opportunities and threats in rail through artificial intelligence", Bhoopathi Rapolu, Infrastructure Intelligence, 16th December 2015.

these systems are acquiring operational expertise that would otherwise require humans with years of experience.

## 2.4 Driverless cars

AI is a critically important technology for self-driving cars. In order to replace a human driver, the system must be able to 'see' all the objects in its immediate vicinity and decide whether it is necessary to apply the brakes. In some emergency situations, the car may have to make a value judgement such as "should I crash into that lamp post in order to avoid crashing into that animal that has just run in front of me?" As in the case of human drivers, such judgements are made on the basis of training and past experience. Tesla collects data from their whole fleet of self-driving vehicles in order to improve 'fleet learning'[3].

AlphaGo has demonstrated that it is possible for machines to become more highly skilled than humans. This suggests that driverless cars could eventually develop to a point where humans are no longer permitted to drive.

## 2.5 Security

Facial recognition is already used in some airports to speed-up passport control. However, facial recognition also works on CCTV images, and this opens up a much wider range of security applications within airports. For example, departing passengers can have their identity checked once when they arrive at the airport, and CCTV-based identification can then be used instead of passport checks as they pass through security and board the aircraft. A similar approach can be used to identify wanted terrorists and criminals.

[3] 'How to implement AI and machine learning', TechRepublic, 2016.

# 3 Risks

As in the case of any powerful technology, AI has enormous potential to do both good and evil. Some of the new risks created by AI are discussed below.

## 3.1 The workings of many AI algorithms are not open to scrutiny

If a doctor or a financial advisor recommends a particular course of action, you expect them to be able to justify their recommendation based on the evidence available and their assessment of the risks. While AI often produces a perfectly valid recommendation, it usually can't provide a logical justification for that recommendation. Large neural networks are enormously complex structures, so even the creators of AI systems sometimes struggle to explain particular results.

However, explainability is key. If AI is to be used to make decisions that significantly affect peoples' lives, then it must be possible to justify those decisions. Nobody is going to accept a self-driving car that can't explain why it sometimes runs over children, or a medical system that inexplicably fails to diagnose a life-threatening illness. For this reason, critical applications often use AI tools paired with humans to optimise the trade-off between human and machine intelligence.

This is one area in which regulation may be catching up with the technology. The EU's General Data Protection Regulation (GDPR), which became effective on 25th May 2018, includes rights relating to automated decision-making and profiling. If someone believes that an organisation is making significant decisions about them based entirely on automated processes, then they can ask for human intervention. There are now discussions within the EU about requiring organisations to provide public justification for the way in which their AI systems are designed[4].

## 3.2 Training data can introduce unintended bias

Latanya Sweeney of Harvard University noticed that her Google search results were accompanied by advertisements asking whether she had ever been arrested. These advertisements did not appear for her white colleagues, and it was eventually discovered that they were being directed towards people whose names suggested that they may be black[5]. In other words, the data used by the AI system had somehow acquired an unintended racial bias.

If the training data contains any form of bias, then an AI system may exhibit that same bias in its decisions. Biases of this type can be extremely subtle and difficult to detect, but they raise obvious legal and ethical issues.

## 3.3 AI-based crime, terrorism and warfare

A report by Cambridge University[6] has warned that the malicious use of AI could be a threat to global stability. The authors expect novel cyber-attacks such as:

[4] "Computer says no: But why?", Chris Edwards, E&T Magazine, May 2018.

[5] "Higher state of mind", Douglas Heaven, New Scientist, 10th August 2013.

[6] "AI is a threat to global stability, warns Cambridge University report", Jack Loughran, eandt.theiet.org, 21st February 2018.

- Highly believable fake videos that impersonate prominent figures to manipulate public opinion
- Automated hacking
- Finely-targeted spam emails using information scraped from social media
- Exploiting the vulnerabilities of AI systems through adversarial examples and data poisoning
- Crashing fleets of autonomous vehicles
- Turning commercial drones into face-targeting missiles
- Holding critical infrastructure to ransom

In the hands of malign individuals or organisations, AI is a powerful weapon that could be used to cause serious harm.

There is now a public debate about the dangers of AI. The late Professor Stephen Hawking warned that AI needs to be controlled or it could do severe damage to humanity. Similarly, Elon Musk (CEO of Tesla and SpaceX) has stated that AI is a fundamental risk to the existence of human civilisation, and called for tougher government regulation. However, this view has been dismissed by Mark Zuckerberg (CEO of Facebook) who prefers to focus on AI benefits such as better disease diagnosis and fewer car crashes.

## 3.4 Impact on jobs

A large number of jobs might disappear in the next 20 years as a result of AI. These include: drivers of vehicles or machinery, farmers, printers and publishers, shop cashiers, travel agents, manufacturing workers, dispatchers (for taxis, ambulances etc), waiters and bartenders, bank tellers, military pilots, soldiers, fast food workers, telemarketers, accountants, stock traders and construction workers[7].

PwC has warned that the increasing use of AI by businesses to replace workers could lead to a 'cliff-edge' scenario in which huge swathes of the working population suddenly lose their jobs[8]. While they suggest that 30% of UK jobs could be at high risk by 2030, they also point out that AI will create new jobs and could drive up productivity and economic growth.

In 1841, 22% of workers were employed in agriculture in England and Wales. By 2013, the proportion was less than 1%. During the same period, employment in manufacturing fell from 36% to 9%. Despite this, unemployment is currently at historically low levels.

While job losses from previous technological advances mainly affected unskilled and semi-skilled workers, AI will also have a significant impact on much more highly-skilled sections of the workforce such as lawyers, accountants and engineers. One scenario envisages "a tiny fraction of machine makers taking almost all the wealth, and low-paid jobs or no work for anyone else"[9]. This could have very serious social consequences.

However, similar predictions were made in the late 1970's as a result of the invention of the microprocessor. It is always easy to see how jobs will be lost as a result of new technologies, but history suggests that new jobs will be created that do not currently exist. Furthermore, automation enhances productivity, so we can expect AI to drive down costs and improve living standards. The 'cliff-edge' predicted by PwC may occur – and would be painful – but we can expect that the employment situation will eventually resolve itself.

[7] '15 jobs that could disappear in the next twenty years as a result of AI', www.alux.com.

[8] 'AI could lead to a cliff-edge scenario of mass unemployment, PWC warns', E&T Magazine, 3 July 2017.

[9] 'If big data is the new oil, is artificial intelligence the new climate change?', Simon Harrison, www.infrastructure-intelligence.com, 23 May 2018.

# 4 Legal & regulatory issues

If some of the more pessimistic predictions about AI turn out to be correct, then the consequences will be very serious. However, the corporations and research organisations that are developing the technology are more likely to be driven by competitive pressures than by any particular concern about global consequences. In situations such as this, national governments and international bodies need to intervene. Decision-making at this level is notoriously difficult, so now is the time to make a start.

## 4.1 Data protection

The EU's General Data Protection Regulation (GDPR) came into force at a time when the use (and misuse) of personal data was very much in the news as a result of the activities of Cambridge Analytica. The GDPR gives EU citizens the right to find out how their data is being used, and to object if they wish.

Time will tell, but it seems likely that GDPR could have a significant impact on internet-based companies that – either directly or indirectly – use AI to produce highly-targeted advertising and promotional material. However, the revenues that can be generated from this data are very large, so companies will not give them up without a fight.

## 4.2 Legal responsibility

On 23 March 2018, the driver of a Tesla electric car was killed in California while the vehicle was operating on Autopilot. The Autopilot feature can change lanes and self-park, but the driver is still required to take control in emergency situations. Earlier in the month, a self-driving Volvo SUV that was being tested by the Uber ride-hailing service hit and killed a pedestrian in Arizona.

Self-driving cars provide a good illustration of the kinds of legal issues that will be encountered as AI develops. Under current laws, driving responsibilities may be shared between a human driver and some form of automation, but the human is ultimately in charge; in the event of an accident, it is the human driver who is legally responsible for any errors.

However, a truly autonomous vehicle would raise some difficult questions:

- If there is a crash, how is legal responsibility assigned between the owner of the car, the manufacturer of the car, the manufacturer of the AI system and the insurance company?
- Would responsibility change if the crash was caused by a cyber-attack on the AI system?
- Would responsibility change if the autonomous vehicle had an override switch that would allow a human to take control?
- Would responsibility change if the occupants of the car were drunk at the time of the crash?

# 5 Conclusions

AI's tremendous benefits must be offset against its potential to cause harm. The industrial revolution in the 18th and 19th centuries gave us massively-improved living standards but left us with a climate change problem that could ultimately destroy us. The development of atomic physics in the 20th century gave us new cures for cancer along with nuclear weapons of unimaginably destructive power. The benefits of AI must not disguise the attendant risks that come with it.

It has been pointed out that "AI is starting to feel uncannily like climate change. Both are initiated by humans, both are definitely happening but their course is unpredictable, and both could drive exponential change in the way we live"[10]. AI enthusiasts argue that humans will never lose control of AI because it is humans that decide when and where the technology should be deployed. However, the same can be said about nuclear weapons.

History has taught us that we cannot hold back the march of science. International treaties and government regulation can help but, as recent attempts to stop nuclear proliferation have shown, they are unlikely to have much impact on rogue states. The problem has to be addressed from a number of different angles:

- AI technologies will certainly be open to misuse, so we need to identify emerging threats. This implies that the technologies should be developed as quickly as possible in strong, stable democracies where they can be adequately supervised.
- Investment is required to establish and maintain defences against AI-based crime, terrorism and warfare. Recent experience with cyber threats has shown that developing effective counter-measures can be a slow and difficult process.
- Open discussion should be stimulated to ensure that public opinion is not left behind by the pace of AI developments. Ethics committees (similar to those used in medicine) should be created to monitor developments and set boundaries.
- Users must be made aware if AI is being used in a product or service. Where AI is being used, the reasons for using AI should be completely transparent.
- Law-makers and regulators must keep pace with the speed of AI developments. Technological progress will not slow down to suit their own more-ponderous ways of working.

[10] 'If big data is the new oil, is artificial intelligence the new climate change?', Simon Harrison, www.infrastructure-intelligence.com, 23 May 2018.

Dr Andrew Wheen is a Project Director within Mott MacDonald's Digital Infrastructure practice. He is part of an integrated team of strategic advisers, technical experts and project managers that are delivering major telecommunications and IT projects around the world.

For further information, contact:

E-mail: Andrew.Wheen@mottmac.com

Office: +44 (0)20 7651 0067

Mott MacDonald Limited, 10 Fleet Place, London, EC4M 7RB, United Kingdom

www.mottmac.com