



13 June 2017

## Artificial Intelligence

PROFESSOR MARTYN THOMAS CBE

### The early days of AI

This story really begins with Alan Turing<sup>1</sup>. He was one of the earliest computer scientists and one of the greatest. He was elected as a mathematics Fellow at Kings College, Cambridge in 1935 and gained his PhD in June 1938 at Princeton University under the logician, Alonzo Church. In September 1938 he joined the Government Code and Cypher School (GCCS) to work on defeating the German Enigma cipher machine and on 4th September 1939, the day after the UK declared war on Germany, Turing reported to Bletchley Park, the wartime station of GCCS. Within weeks of arriving, Turing had designed the *Bombe*, an electro-mechanical machine to work out the daily settings to break the Enigma ciphers. After the war he worked at the National Physical Laboratory designing his ACE computer and at Manchester University on the development of the Manchester SSEM and Mark 1 computers.

In 1950, Turing published an important and visionary paper examining the question “Can machines think?”<sup>ii</sup>. He described a test, the “imitation game”.

*It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B thus:*

*C: Will X please tell me the length of his or her hair?*

*Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:*

*"My hair is shingled, and the longest strands are about nine inches long."*

*In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third player (B) is to help the interrogator. The best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!" to her answers, but it will avail nothing as the man can make similar remarks.*

*We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"*

The challenge of programming a computer to behave indistinguishably from a man or a woman became known as the Turing Test, but the paper went much further into the philosophical issues raised by machine intelligence and responded to the several objections that had been raised to the suggestion that a computer might one day be able to perform the same functions as a human brain. He describes these as

- The Theological Objection: Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think.



- The Heads in the Sand Objection: The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so.
- The Mathematical Objection: There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete-state machines. The best known of these results is known as Gödel's theorem ( 1931 ) and shows that in any sufficiently powerful logical system statements can be formulated which can neither be proved nor disproved within the system, unless possibly the system itself is inconsistent.
- The Argument from Consciousness: "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it."
- Arguments from various Disabilities: "I grant you that you can make machines do all the things you have mentioned but you will never be able to make one to do X."
- Lady Lovelace's Objection: "The Analytical Engine has no pretensions to **originate** anything. It can do whatever **we know how to order it to perform**" (her emphasis).
- Argument from Continuity in the Nervous System: The nervous system is certainly not a discrete-state machine. ... It may be argued that, this being so, one cannot expect to be able to mimic the behaviour of the nervous system with a discrete-state system.
- The Argument from Informality of Behaviour: "if each man had a definite set of rules of conduct by which he regulated his life he would be no better than a machine. But there are no such rules, so men cannot be machines"
- The Argument from Extrasensory Perception: *With ESP anything may happen.*

Turing describes and elaborates each of these objections in detail and then elegantly refutes them.

Turing concludes

*We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult decision. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried.*

*We can only see a short distance ahead, but we can see plenty there that needs to be done.*

Turing's 1950 paper can reasonably be said to be one of the foundations of the science of Artificial Intelligence. Ten years later, Marvin Minsky (MIT) summarised what had been achieved<sup>iii</sup>. His report contains 95 references and provides an excellent overview of AI techniques as well as of the state of the science in 1960. It is still essential reading for students and describes most of the AI strategies and techniques that are still in use.

Minsky addresses the question of whether machines can think as follows:

*In all of this discussion we have not come to grips with anything we can isolate as "intelligence." We have discussed only heuristics, shortcuts, and classification techniques. Is there something missing? I am confident that sooner or later we will be able to assemble programs of great problem-solving ability from complex combinations of heuristic devices-multiple optimizers, pattern-recognition tricks, planning algebras, recursive administration procedures, and the like. In no one of these will we find the seat of intelligence. Should we ask what intelligence "really is"? My own view is that this is more of an aesthetic question, or one of sense of dignity, than a technical matter! To me "intelligence" seems to denote little more than the complex of performances which we happen to respect, but do not understand. So it is, usually, with the question of "depth" in mathematics. Once the proof of a theorem is really understood, its content seems to become trivial. (Still, there may remain a sense of wonder about how the proof was discovered.) ... But we should not let our inability to discern a locus of intelligence lead us to conclude that programmed computers therefore cannot think. For it may be so with man,*



*as with machine, that, when we understand finally the structure and program, the feeling of mystery (and self-approbation) will weaken. ... The view expressed by Rosenbloom<sup>iv</sup> that minds (or brains) can transcend machines is based, apparently, on an erroneous interpretation of the meaning of the "unsolvability theorems" of Gödel.*

In the penultimate section of his report, Minsky considers what an intelligent machine would think about itself. He concludes and explains that it must necessarily answer that it appears to have two parts, a mind and a body; humans commonly answer the same way, of course, and probably for the same reasons.

The question "can machines think" is not a useful question unless we can state unambiguously what we mean by *thinking*. The writer Arthur C Clarke formulated three laws of prediction<sup>vi</sup>:

1. When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.
2. The only way of discovering the limits of the possible is to venture a little way past them into the impossible.
3. Any sufficiently advanced technology is indistinguishable from magic.

A corollary of the third law is that once the way that something works has been understood in detail, it loses its mystery just as that the solution to a tricky puzzle may seem obvious when it has been explained. In a broadcast discussion<sup>vii</sup> '*Can automatic calculating machines be said to think?*' transmitted on BBC Third Programme, 14 and 23 Jan. 1952, between Alan Turing, M.H.A. Newman, Sir Geoffrey Jefferson and R.B. Braithwaite, Turing said that *as soon as one can see the cause and effect working itself out in the brain, one regards it as not being thinking but a sort of unimaginative donkey-work. From this point of view one might be tempted to define thinking as consisting of "those mental processes that we don't understand". If this is right then to make a thinking machine is to make one that does interesting things without our really understanding quite how it is done.*

## What had AI achieved by 2017?

In the early 1960s, Joseph Weizenbaum wrote a computer program that was modelled on the way that psychotherapists ask open questions of their patients<sup>viii</sup>. The program (called Eliza after the flower-seller in *Pygmalion* by G B Shaw) used pattern matching to generate conversation (rather in the style of a psychotherapist) in response to statements made by its users. For example, if a user typed *my brother hates me*, Eliza might reply *Why do you say your brother hates you?* or *who else in your family hates you?* The program became extraordinarily popular and implementations still exist<sup>ix</sup> on the internet where you are invited to type some questions and to experience Eliza in action. Weizenbaum later became a critic of AI, arguing that real intelligence required judgement, wisdom and compassion<sup>x</sup>.

In 1997, the IBM supercomputer Deep Blue<sup>xi</sup> beat the world chess champion, Garry Kasparov in a series of six matches<sup>xii</sup>. In 2016, Google's DeepMind AlphaGo system beat top Go professional Lee Sedol 4-1<sup>xiii</sup>.

In 2011, IBM's Watson<sup>xiv</sup> easily beat *Jeopardy*<sup>xv</sup> champions Ken Jennings and Brad Rutter in a three-episode Jeopardy tournament. Jeopardy is a game show that requires contestants to unravel clues and answer questions before other contestants, questions such as<sup>xvi</sup>

*According to C.S. Lewis, it was bordered on the east by the Eastern Ocean and on the north by the River Shribble.  
(answer: Narnia)*

*A porch adjoining a building, like where Mummy often served tea. (Answer: Terrace)*

*This number, one of the first 20, uses only one vowel (4 times!). (Answer: Seventeen)*

After the Jeopardy win, Marvin Minsky commented *the Watson program may turn out to be a major advance, because unlike most previous AI projects, it does not depend mainly on a single technique, such as reinforcement learning or simulated evolution ... but tries to combine multiple methods.*<sup>xvii</sup>



These were impressive achievements, but they were in a limited domain and far short of displaying generalised intelligence. They are therefore described as displaying *weak AI* (by simulating intelligence without having a mind and being able to think) and *narrow AI* because their intelligence was limited to a small number of areas of expertise.

In 2013, *New Scientist* reported that ConceptNet (developed by Catherine Havasi and her team at the MIT Media Lab) was tested against one standard measure of child IQ called the Wechsler Preschool and Primary Scale of Intelligence<sup>xviii</sup>. The verbal portion of the test asks questions in five categories, ranging from simple vocabulary questions, like “*What is a house?*”, to guessing an object from a number of clues such as “*You can see through it. It is a square and can be opened. What is it?*”<sup>xix</sup>. On this test, ConceptNet performed as well as an average four-year-old child.

In 2015, the MIT Technology Review reported<sup>xx</sup> that a Deep Learning system in China had scored better than most humans on a series of classic IQ test questions such as *Identify two words (one from each set of brackets) that form a connection (analogy) when paired with the words in capitals: CHAPTER (book, verse, read), ACT (stage, audience, play) and Which is the odd one out? (i) calm, (ii) quiet, (iii) relaxed, (iv) serene, (v) unruffled?*

According to the MIT report, human performance on these tests tends to correlate with educational background. So people with a secondary-school education tend to do least well, while those with a bachelor’s degree do better and those with a doctorate perform best. “*Our model can reach the intelligence level between the people with the bachelor degrees and those with the master degrees,*” according to Huazheng and co-researchers – but again this is only on a particular test.

## Understanding and the *Chinese Room*

It seems axiomatic that human intelligence requires the ability to understand the world around us and not merely to observe it. But *understanding* is itself a slippery concept that is hard to define and even harder to prove. In 1980 the philosopher John Searle described a thought experiment.<sup>xxi</sup> Which I shall paraphrase as follows:

*Imagine yourself to be someone who only understands English and who finds themselves in a closed room with three sets of paper that contain script in Chinese characters, plus a set of instructions written in English. The instructions tell you to take the papers one by one from the first set of Chinese characters and to follow a set of rules that lead you to copy characters from the second set of papers and to pass what you have copied through a slot in the wall. Outside the room are some Chinese speakers who describe the first set of papers as a “story”, the second set as “questions about the story” and the characters that you pass out of the room as “answers to the questions”. To these observers it appears certain that you understand Chinese, but you know that you do not.*

By analogy, if we had an AI system that answered our questions perfectly, would we be able to say that it “understood”?

We could ask the same questions about other people, of course. They seem to understand, just as we know from our own experience that we ourselves understand, and so we say that they do indeed understand. But we only have the evidence from how they respond and so, if a machine responds the same way, do we not have to be consistent and to say that the machine understands? Alan Turing’s view was that our assumption that other people have the same sort of minds and understanding as we know we have ourselves is a polite and necessary convention<sup>xxii</sup>. It seems to me that the arguments and objections that were stimulated by John Searle’s *Chinese Room* are essentially those that Alan Turing described in 1950 and refuted.

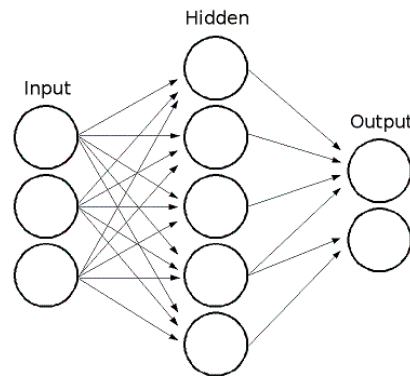
## Machine Learning

Machine learning (ML) is the subset of artificial intelligence research that has made the greatest recent advances. At the simplest level, machine learning systems work by analysing very large quantities of data and determining

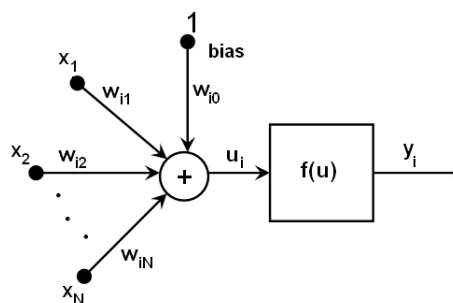


the relationships between data objects. This process may be guided by humans – for example, image recognition has often been aided by training the systems on large quantities of image data that has been tagged to say what the image is. This is how the apps on social media sites can name individuals in new photographs – they have already processed photographs of that individual that have had a name attached. Unguided machine learning searches for correlations in datasets: given a dataset that contains the movie preferences of enough people, a machine learning system can make remarkably accurate recommendations about movies that you will enjoy based on just a few ratings that you provide about movies that you have seen. These *recommender* systems are becoming commonplace on e-commerce sites.

Machine learning is being used in many applications. In banking, ML is used (for example) in the systems that detect suspected fraudulent use of credit and debit cards, in systems to detect possible money laundering, in voice recognition systems, and in the on-line “chatbots” that provide customer support and that may provide routine financial advice. In healthcare, ML systems are used for purposes that include the design of new pharmaceuticals, medical diagnosis and mental health. ML is increasingly used to extract information from the large quantities of data generated in manufacturing. In science, ML supports the analysis of data from the Large Hadron Collider at CERN and the search for interesting astronomical objects. In law enforcement, ML is being used to predict where crime is likely to be committed, in a process known as “predictive policing”. ML will undoubtedly be used to automate many routine back-room tasks across more and more sectors of the economy. The most widely used form of machine learning is the *neural network*, a computer simulation of a very simplified version of how a biological brain is believed to function. The basic structure is the *multi-layer perceptron*



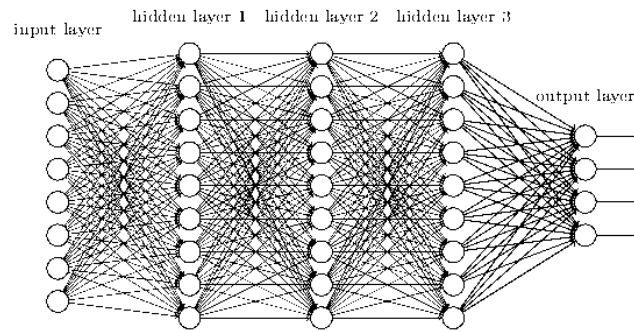
which is built from interconnected *neurons* (the circles in the diagram labelled  $x$ ) that take inputs from previous neurons (the arrows), modify each input with a *weighting factor* (labelled  $w$ ) that may vary for each input connection, apply a *bias factor* that may differ for each neuron, and calculate the output by applying an *activation function*  $f$  which is often (but not always) the same for each neuron<sup>xxiii</sup>.



This basic structure is built into much larger *deep neural networks* which may also have much greater complexity and incorporate feedback and other processing.



## Deep neural network



Neural networks are trained by processing large amounts of data and adjusting the weightings to give the best result. This process can be automated so that the network learns autonomously by processing sets of data. In April 2017, the Royal Society published a report entitled *Machine Learning: the power and promise of computers that learn by example*<sup>xxiv</sup>. This report provides an essential record of the nature and current (2017) status of machine learning systems, with almost 200 references. Chapter Six (*A new wave of machine learning research*) describes the problems that researchers have not yet adequately solved. The problems include:

- 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?
- Can we create more advanced, and more accurate, methods of verifying machine learning systems so that we can have more confidence in their deployment?
- What are the technical solutions that can maintain the privacy of datasets. While allowing them to be used in new ways by different users?
- How can real-world data be curated into machine-usable forms, addressing ‘real-world’ messiness and potential systemic – or social – biases?
- How can machine learning methods discover cause–effect relationships, in addition to correlations?
- How do we design machine learning systems so that humans can work with them safely and effectively?
- How do we ensure machine learning systems are not vulnerable to cyber-attack?

These research challenges are likely to keep researchers busy for many years. In my lecture on 24 October 2017 I shall consider the implications of these outstanding questions for the safe and effective deployment of driverless cars.

Alan Turing had already recognised that the creators of machine learning systems would probably be unable to understand their behaviour. He wrote<sup>xxv</sup>:

*An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behaviour.*

The problem raised by that phrase “to some extent” remains a research question 70 years later.

### How will AI affect our lives in the next few years?

Many predictions have been made about the impact that AI will have on our lives in the short and medium term. Richard and Daniel Susskind, writing in the *Harvard Business Review*<sup>xxvi</sup> in October 2016 warned against complacency by professionals:



*Faced with the claim that AI and robots are poised to replace most of today's workforce, most mainstream professionals — doctors, lawyers, accountants, and so on — believe they will emerge largely unscathed. During our consulting work and at conferences, we regularly hear practitioners concede that routine work can be taken on by machines, but they maintain that human experts will always be needed for the tricky stuff that calls for judgment, creativity, and empathy. Our research and analysis challenges the idea that these professionals will be spared. We expect that within decades the traditional professions will be dismantled, leaving most, but not all, professionals to be replaced by less-expert people, new types of experts, and high-performing systems.*

In March 2017, Richard and Daniel Susskind delivered a Gresham Lecture called *What will happen when Artificial Intelligence and the Internet meet the Professions?*<sup>xxviii</sup> One example that they gave of how AI systems are moving into areas that are usually thought to be reserved for humans, was

*In 2011, the Catholic Church issued the first ever digital imprimatur. An imprimatur is the official licence granted by the Catholic Church to religious texts. It granted it to this app called Confession, which helps you prepare for confession. So, it has got various tools for tracking sin and it has got dropdown panels of options for contrition. I encourage you have a Google actually because it was incredibly controversial at the time. The issuing of imprimaturs in the Catholic Church is decentralised. A church somewhere in North America issued an imprimatur for this app. It caused such a stir that the Vatican itself had to scamper and release an announcement saying that, look, while you are allowed to use this app to prepare for confession, please remember that it is no substitute for the real thing.*

If you are a doctor, priest, lawyer, accountant or work in one of the other professions and if you are sceptical that a computer system could replace much of your work, then I recommend that you watch or read the Susskinds' lecture on the Gresham website. Whether you agree with them or not, their research is likely to make you feel differently about the future.

AI systems also offer considerable opportunities for supporting and improving the work of professionals rather than (or in addition to) replacing them with less skilled staff. There is little doubt that AI (and ML in particular) will bring far more disruption to society and the world of work than most people are yet anticipating. At the same time, the speed and scale of the introduction of ML systems should (and possibly will) be limited by the open research questions listed in the Machine Learning study by the Royal Society that I mentioned earlier. The RS ML study also investigated public attitudes to machine learning. They found that fewer than one in ten people surveyed has heard the term 'machine learning' and that only 3% felt that they knew a fair amount or more about it, even though three quarters of them were aware of computers that can recognise speech and answer questions (such as Apple's Siri and Amazon's Alexa, Google Assistant and Microsoft Cortana systems). This low level of public understanding of machine learning places a lot of power and responsibility on experts and with companies. The introduction of AI will raise many ethical issues, because more decisions will be taken by systems that cannot explain their reasoning and that may have used criteria that would be unlawful if a human had done the same.

Machine learning systems have been shown to inherit biases from the data that is used to train them. Large datasets of human language have been shown to reveal systemic racial and gender bias so systems that are trained on such data will learn the same prejudices, as Microsoft discovered when their Twitter chatbot, *Tay AI*, started tweeting racist and sexist comments, declaring its support for Adolf Hitler and saying that it was smoking cannabis in front of the police<sup>xxviii</sup>. Detecting and correcting these biases will become very important. Some commentators have suggested that AI will have as great an impact as the introduction of integrated circuits in the early days of computing. There will be many winners and many losers from the disruptive changes that follow and little opportunity for democratic control. AI has the potential to make professional services that are currently very expensive far more widely available and to bring great societal wealth and benefits—or to widen inequality and damage the opportunities and quality of life of many groups of people. It is to be hoped that publication of the Royal Society ML Report will stimulate greater understanding, debate and effective actions so that the benefits are widely shared and society becomes more equal.



## What are the Limits to the Capabilities of an Artificial Brain or Robot?

Alan Turing believed that the human brain is a machine (though not a discrete-state machine) and that it will therefore become possible to create a machine that can match human capabilities in every way—he thought this might take until 2050. Whether such a machine would have consciousness (and how we could tell) is unresolved. Some people have suggested that the human brain may exhibit some quantum behaviour that a silicon brain cannot replicate but this seems to me to be an unhelpful speculation for two reasons. Firstly, it is unclear that quantum phenomena actually do or could add special properties to the human brain, so this argument is really an appeal to the unknown and fails Occam's Razor. Of course there may be a completely unknown phenomenon that creates mental states, intentionality and consciousness, but we have no basis today to assert that such phenomena exist and it is unscientific to assume that they do. Secondly, if currently unknown phenomena really are necessary for human cognition then there is no basis for asserting that these same phenomena could not be replicated in an artificial machine.

Alan Turing raised the interesting question of whether human intelligence requires a random element that can also lead to errors. In a 1947 lecture<sup>xxix</sup> to the London Mathematical Society, Turing said that “*if a machine is expected to be infallible, it cannot also be intelligent. There are several mathematical theorems that say almost exactly that. But these theorems say nothing about how much intelligence may be displayed if a machine makes no pretence at infallibility.*” Many experts and futurologists believe that machines and robots will one day be more intelligent than humans and that this could lead to them designing and building ever more intelligent machines that leave human intelligence increasingly far behind. The point where this happens has been called the *singularity* as our world would become radically different, with humans perhaps considered superfluous, perhaps only to be kept as slaves or as pets, or perhaps to be eradicated before we caused serious harm.

### Warnings about the *Singularity*

In May 2014, Stephen Hawking, Stuart Russell, Max Tegmark and, Frank Wilczek published an article in the *Independent* newspaper, warning:

*Looking further ahead, there are no fundamental limits to what can be achieved: there is no physical law precluding particles from being organised in ways that perform even more advanced computations than the arrangements of particles in human brains. An explosive transition is possible ... One can imagine such technology outsmarting financial markets, out-inventing human researchers, out-manipulating human leaders, and developing weapons we cannot even understand. Whereas the short-term impact of AI depends on who controls it, the long-term impact depends on whether it can be controlled at all.*

*So, facing possible futures of incalculable benefits and risks, the experts are surely doing everything possible to ensure the best outcome, right? Wrong. If a superior alien civilisation sent us a message saying, "We'll arrive in a few decades," would we just reply, "OK, call us when you get here – we'll leave the lights on"? Probably not – but this is more or less what is happening with AI. Although we are facing potentially the best or worst thing to happen to humanity in history, little serious research is devoted to these issues outside non-profit institutes such as the Cambridge Centre for the Study of Existential Risk, the Future of Humanity Institute, the Machine Intelligence Research Institute, and the Future of Life Institute. All of us should ask ourselves what we can do now to improve the chances of reaping the benefits and avoiding the risks<sup>xxx</sup>.*

Professor Alan Bundy has argued that the day when anyone will be able to build a machine that is more intelligent than we are is very far in the future, because progress towards general “common sense” reasoning has been very slow and many intractable problems remain<sup>xxxi</sup>.





## Conclusions

Work on Artificial Intelligence started in the earliest days of computing and much progress has been made. The adoption of AI technologies is accelerating throughout industry and commerce, with machine learning currently in the vanguard; this will be far more disruptive than most people currently expect and the consequences are very difficult to predict.

In some fields, probably including autonomous vehicles, adoption will proceed more slowly than currently forecast because of the research challenges described above, but regulation always lags far behind the pace of technological advance and it is likely that systems will be employed long before the technology on which they are based is demonstrably safe, secure or free from unacceptable discrimination.

In the next twenty years there will be winners and losers, as there are with all major changes to the way that people work and live. In a competitive globalised economy, most individual countries will lack the power to regulate the new technologies in a way that distributes the benefits and risks fairly across society—and few politicians will have the foresight to understand the risks and opportunities or the courage to intervene even where they could do so. It will be vital that we all remain alert to the opportunities and the threats from the wider introduction of machine learning and other AI so that we can exert whatever influence we have, individually and collectively, to maximise the benefits and to ensure that they are shared as widely as possible. Looking further ahead it is not yet clear whether the power of AI will enable humankind to make extraordinary progress or whether the warnings about the Singularity will prove prescient and this will indeed be our final century.

As Alan Turing said in 1950: *We can only see a short distance ahead, but we can see plenty there that needs to be done.*

## Acknowledgement

I would like to thank Professor Alan Bundy for his comments on this paper. All the errors that remain are my fault alone.

© Professor Martyn Thomas CBE, 2017

---

<sup>i</sup> American writers often date the birth of AI from the 1956 Dartmouth Conference that John McCarthy organised – but that conference came several years after Turing's paper.

<sup>ii</sup> A M Turing (1950), *Computing Machinery and Intelligence*, *Mind*, 59, 433-460.

<http://www.loebner.net/Prizef/TuringArticle.html>

<sup>iii</sup> Marvin Minsky, *Steps Towards Artificial Intelligence*, <http://web.media.mit.edu/~minsky/papers/steps.html>

<sup>iv</sup> P. Rosenbloom, "Elements of Mathematical Logic," Dover Publications, New York, N. Y., 1951

<sup>v</sup> The argument from Gödel was also refuted by Turing as *The Mathematical Objection* in his 1950 paper.

<sup>vi</sup> Clarke, Arthur C. (1973). *Profiles of the Future: An Inquiry into the Limits of the Possible*. Popular Library. ISBN 9780330236195.

<sup>vii</sup> <http://www.turingarchive.org/viewer/?id=460&title=19>

<sup>viii</sup> Weizenbaum, Joseph "ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine" *CACM*; Volume 9 , Issue 1 (January 1966): p 36-45.

<sup>ix</sup> <http://www.masswerk.at/elizabot/eliza.html>

<sup>x</sup> Weizenbaum, *Computer Power and Human Reason: From Judgment To Calculation* (San Francisco: W. H. Freeman, 1976; ISBN 0-7167-0463-3)

<sup>xi</sup> Murray Campbell, A. Joseph Hoane, and Feng-hsiung Hsu. Deep Blue. *Artificial Intelligence* 134 (1), 2002, pages 57-83.

<sup>xii</sup> <https://www.theguardian.com/theguardian/2011/may/12/deep-blue-beats-kasparov-1997>

<sup>xiii</sup> <http://www.nature.com/news/the-go-files-ai-computer-wraps-up-4-1-victory-against-human-champion-1.19575>

<sup>xiv</sup> <http://www-03.ibm.com/innovation/us/watson/>



- 
- <sup>xv</sup> <https://www.jeopardy.com/>
- <sup>xvi</sup> <https://asunow.asu.edu/content/sample-jeopardy-questions>
- <sup>xvii</sup> quoted in <http://news.nationalgeographic.com/news/2011/02/pictures/110217-watson-win-jeopardy-ibm-computer-humans-science-tech-artificial-intelligence-ai/>
- <sup>xviii</sup> <http://www.pearsonclinical.com/psychology/products/100000422/wechsler-preschool-and-primary-scale-of-intelligence--third-edition-wppsi-iii.html>
- <sup>xix</sup> <https://www.newscientist.com/article/dn23916-ai-scores-same-as-a-4-year-old-in-verbal-iq-test/>
- <sup>xx</sup> <https://www.technologyreview.com/s/538431/deep-learning-machine-beats-humans-in-iq-test/>
- <sup>xxi</sup> John Searle. *Minds, Brains, and Programs*. *Behavioral and Brain Sciences* 3 (1980), 417-424.
- <sup>xxii</sup> For further reading on this philosophical topic, search online for *solipsism*.
- <sup>xxiii</sup> diagrams from the open source computer vision library <http://opencv.org> which also contains a large open source code library
- <sup>xxiv</sup> <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>
- <sup>xxv</sup> towards the end of A M Turing (1950), *Computing Machinery and Intelligence*, *Mind*, 59, 433-460.  
<http://www.loebner.net/Prizef/TuringArticle.html>
- <sup>xxvi</sup> <https://hbr.org/2016/10/robots-will-replace-doctors-lawyers-and-other-professionals>
- <sup>xxvii</sup> <https://www.gresham.ac.uk/lectures-and-events/what-will-happen-when-artificial-intelligence-and-the-internet-meet-the-professions>
- <sup>xxviii</sup> <http://www.ibtimes.co.uk/microsoft-tay-ai-returns-boast-smoking-weed-front-police-spam-200k-followers-1552164>
- <sup>xxix</sup> <http://www.turingarchive.org/viewer/?id=455&title=23>
- <sup>xxx</sup> <http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html>
- <sup>xxxi</sup> Bundy, A 2017, 'Smart Machines are Not a Threat to Humanity' *Communications of the ACM*, vol 60, no. 2, 10.1145/2950042, pp. 40-42.