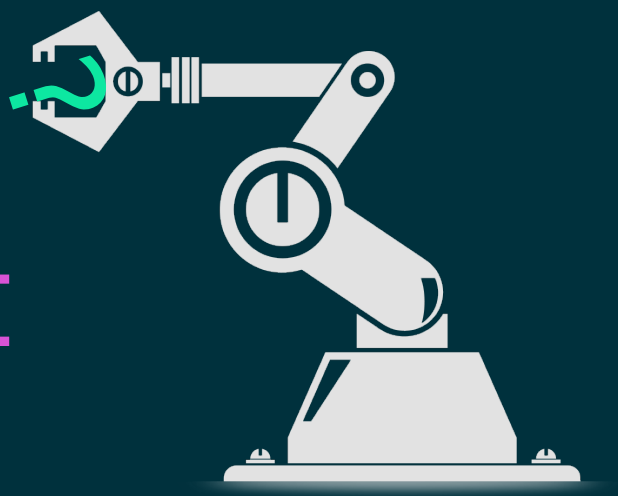
THE CYBER SOLICITOR

# Will Humans Need Not Apply



*What improvements in AI could do to the future economy and greater humanity*

June 2017

"The lives and experiences of all other animals were undervalued, because they fulfilled far less important functions, and whenever an animal ceased to fulfil any function at all, it went extinct. However, once humans lose their functional importance to the network, we will discover that we are not the apex of creation after all. The yardsticks that we ourselves have enshrined will condemn us to join the mammoths and the Chinese river dolphins in oblivion. Looking back, humanity will turn out to be just a ripple within the cosmic data flow."

*Homo Deus: A Brief History of Tomorrow*, by Yuval Noah Harari

1

# Brief

June 2017

Topic | AI & Robotics

## Will Humans Need Not Apply?

It has taken some time, but now artificial intelligence has become one of the most illustrious technologies of today which has begun to shake up the world in an unprecedented fashion. What has been made it even more fascinating is its potential, as its applicability continues to broaden and encompass a range of industries and ways of life, possibly changing them in very significant ways. As a result, artificial intelligence could drastically change the way that the economy works, as well as having impacts on the traditional conventions of society. With that will also come interesting and probably difficult questions regarding the legal and ethical implications of the rise of artificial intelligence. As such, there is a great importance in ensuring that the right decisions are made with regard to AI in order for its rise to be compatible with what is best for humanity.

THE CYBER SOLICITOR

Vol No. 1

Issue No. 2

# What improvements in AI could do to the future economy and greater humanity

Back in 1997, a computer was able to defeat a human being at a game of chess. The victor, a machine called Deep Blue, managed to beat Garry Kasparov, one of the best human players of all time. The chess-playing computer was developed by IBM, written in C (a computer programming language), and thus contained enough computing power to evaluate 200 million positions per second. It used a technique known as brute-force to evaluate these millions of positions and to choose the best moves to defeat Mr. Kasparov.

But fast-forward to March of 2016, AlphaGo, a computer program developed by researchers at the London-based company DeepMind, which is now owned by Google, won a five-game match of Go against Lee Sedol, considered the game's best player. Developing computers which can successfully defeat human opponents at this ancient East Asian game is a much more complex task than one which can play and beat a human at chess. The estimated number of possible positions of stones on a Go board is $10^{170}$.

This shows just how much computers have managed to improve over the past 20 years or so. It is not just the technologists computer scientists who are excited; economists, philosophers, politicians, lawyers and many others will take great interest in the latest developments taking place in AI (artificial intelligence), a technology allowing computers to do tasks which could traditionally only be done by humans. In addition, there are other technologies that are helping push these significant shifts. Machine learning, a way of achieving artificial intelligence, is the practice of analysing data to make determinations or predictions, essentially 'training' computers to complete specified tasks. Deep learning is a way to implement machine learning which has now been made possible, thanks to improved computational power and the availability of big data.

Though with the rise of AI, opinions are split. There are some experts who predict that it will have noticeable impacts sooner rather than later, whereas others believe it will still take some time before AI begins to have truly drastic impacts on the world. The impact it may have on employment is an example of where there appears to be a divide, particularly about whether future improvements in AI will leave masses of people unemployed as more advanced machines start to become the preferred choice of capital.

**The Big Questions**

With all the developments taking place in AI and other technologies, the question of where humans fit in all of this remains somewhat of a mystery. Will improvements in technology eliminate the need for humans? If so, to what extent and in which areas in particular? Moreover, what economic, legal and societal impacts could arise as a result? Are such impacts even likely to happen?

Advancements in AI and other related technologies, whether it causes the efforts of humans to become void or not, is likely to have significant impacts on a number of key areas. To start with, the issue of employment is even now a very hot topic with regard to AI, with improvements in technology threatening jobs in numerous professions and industries, even traditionally prestigious areas such as law and medicine. Other areas like cybersecurity and warfare are also likely to experience drastic changes brought upon by improvement in AI.

Also, in addition to the economical impacts, the societal impacts are also likely to arise. Improvements in technology have encouraged a greater reliance on computers and Big Data. There are already plenty of examples of this; Amazon's suggested product list personalised for online users is based on data from shopping history on the site. Google's auto-reply feature in its Gmail application is also a technology based on the use of AI. These continuing improvements in AI, machine learning and Big Data may encourage a culture of 'datatism,' in which humanity becomes increasingly integrated with and reliant upon technology in a way that could diminish our internal authority as it is replaced by the wisdom of machines processing masses of data.

Additionally, the legal implications of improvements in AI will be significant. Who will become liable in an accident involving self-driving cars? Who is held accountable when an AI enabled machine deploys a cyber attack against another country? If robots and AI entities are to be held liable, do they deserve certain rights and liberties too? The rise of AI begs interesting questions, many of which do not have any clear answers yet.

For more Special Reports and analysis, visit thecybersolicitor.com

**Bigger, Better and Faster**

The improvements in AI are beginning to grab more headlines and attention. Yet, this has only been a fairly recent surge. The idea of inanimate objects having the capacity to be intelligent resembling that of human beings has been around for a long time. Such fantasies were even a part of Greek mythology.

Yet, the concept of computational artificial intelligence did not officially develop into its own field until the 20th century. Progress in mathematical logic not only produced the foundations for the first modern computers to come into being, but also for the possibility of building an artificial brain. In the 1950s, scientists from an array of fields, including mathematics, engineering and even

economics, commenced the discussion of creating an artificial brain, which would later blossom into an academic discipline in 1956. It did so at a Conference at Dartmouth College in New Hampshire, where the now vastly popular term 'artificial intelligence' was born.

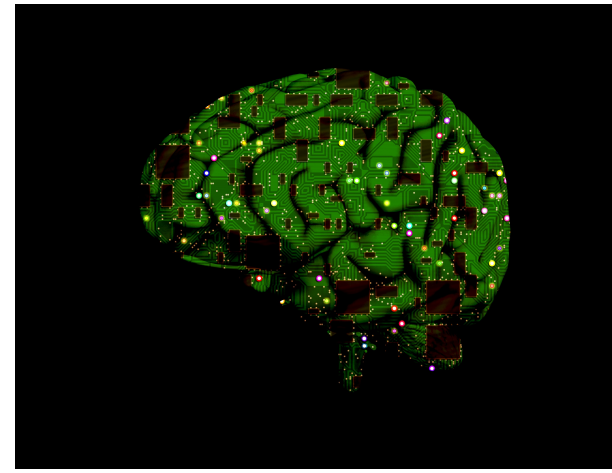
But for a while, the field of AI seemed to not live up to all that it promised. It has taken a number of years for it to improve enough for many more economists, psychologists and other professionals to consider it seriously. But with the help of improvements in other technologies over time, AI became more sophisticated and has now exploded into such a popular area. This is evident in how several tech giants, from Google to Apple, are now looking for talented individuals in the field to advance their own projects, notably self-driving cars.

The improvements in AI which has caused its rise to fame is particularly due to developments in three key areas of computer science; big data, better algorithms and better (and cheaper) computing power.

The availability of high volumes of information generated through several different digital means enables companies, governments and other organisations to observe interesting insights of which, perhaps, could not have identified before. Consider online shopping for example. Booksellers can observe what the bestsellers are and even tie purchases to individual customers through loyalty programs or other schemes. But with online shopping, a whole array of information is suddenly available. Retailers online can now see not only what customers bought, but also what they looked at, which promotions they were most influenced by, which genres they preferred, how much time they spent looking through the site before making a purchase. Retailers can then use this information to produced advertisements and promotions tailored to specific individuals or groups. These vast insights produced by big data are exactly what online retail giant Amazon relies on to sell its products.

For AI, masses of collected data provide machines with the means to learn and execute certain tasks. This is how Alpha Go was able to successfully defeat Go champion Lee Sedol; by sifting through a massive catalogue of data on all the possible moves the machine could execute. Just like human brains rely on past experiences and examples to encounter and process new scenarios, machines are able to do the same with the use of big data.

As well as big data, advancements in algorithms have also enabled machines to essentially learn how to perform tasks independently. This is due to the research in artificial neural nets since the 1950s and has played a major role in today's drastic improvements in AI. Just like a biological brain, layers of artificial neurones process information to produce particular outputs. It takes several of these layers, with millions of neurones, to recognise a human face, for example. Critically, Geoff Hinton of the University of Toronto was able to mathematically optimise these neural processes, a tweak now known as deep learning, which is an important part of Google's search engine and has the potential to be used for a wide variety of other uses. "What got people excited about this field is that in learning technique, deep learning, can be applied to so many different domains, says John Giannandrea, who is the head of machine intelligence at Google.



But to build a neural network for AI software requires plenty of computing power. The introduction of GPU chips (graphics processing unit), originally used to advance video games, allowed neural networks to connect hundreds of millions of nodes and has proven instrumental in the development of AI. Several GPUs running neural networks is how Netflix can make reliable recommendations to its subscribers.

Replicating the capabilities of the human brain in a digital format has proven difficult. But years of research has now given the field of AI a new outlook. Plenty of companies, even some outside of Silicon Valley, are looking to implement AI into their products and services. Further advancements are on the horizon, as the fantasy of computers being able to perform human tasks is starting to come true. If so, what will the potential impacts be?

**Computer Capital**

The fear of machines replacing humans in the workplace is not necessarily a new one. Even in the 18th and 19th centuries, technological revolutions caused great economic change, particularly in the advanced countries.

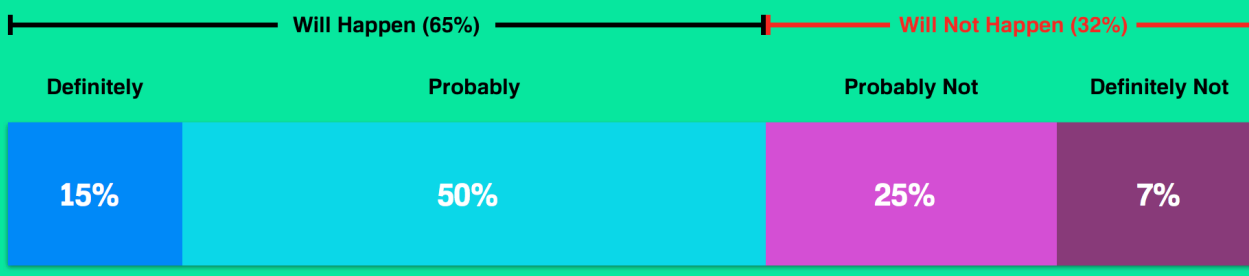
History may indeed repeat itself; the notion that computers may yet again cause shifts in job markets has now come to the fore with the advancement of AI. The main difference between previous waves of computer-induced shake-ups and the one which the world may soon face is that AI is allowing computers to do jobs which traditionally many thought could only be done by humans. The improvements in AI have been so great that even the most prestigious of professions are under threat.

Accordingly, there appears to be a much wider range of jobs and occupations that are particularly vulnerable to the rise of AI. To begin with, jobs and other types of work that involve repetitive routines may be among the first to be affected. Throughout human history, automation has taken over many of these repetitive tasks as it becomes cheaper and more productive. Secondly, those industries and professions which have for a long while managed to resist great changes from technological influences will also be prime targets. The legal profession is a good example, as it is an industry which has mostly remained the same for a long time and has thus, far avoided any mass disruption from improvements in technology. This could change quite soon.

1

## Rise of the Robots

Percentage of adults who believe that computers and robots will takeover much of the work currently done by humans in the next 50 years




Source: Pew Research Centre

Those jobs which feature repetitive routines are ones which can most easily be replaced by computers. For example, a manufacturer worker on a production line whose role features very specific tasks and routines that have to be repeated again and again is at high risk of being automated. This was found to be the case by Carl Benedikt Frey and Michael Osborne, who in 2013 looked at the 702 occupations which could become automated, and found that 47% of workers in America were in jobs that could be subject to automation. Those in logistics and office support roles were amongst the most likely to be replaced by "computer capital." Technology has come far enough to replace humans in these kinds of tasks very easily.

Even jobs in retail, which may not always consist of lots of repetitive tasks, are also at risk. This is because the importance of 'peoples skills' may soon become void with the availability of big data which can allow computers and machines to make much more accurate predictions on what particular customers may want to buy based on their past purchasing habits. In the same way that Deep Blue used masses of information to decide which moves to use in a chess match, machines can use past purchasing data to offer products to customers in a much more efficient manner than any human could possibly manage, and thus, may be able to sell more.

This points to the important economic reasoning for the use of AI and machines; it improves efficiency and increases productivity at a lower cost. An artificially intelligent machine does not need to be paid. More than that, it does not need regular breaks, or health insurance, and is not limited to only being able to work for a certain number of hours or a certain number of days. Due to this, computers can work almost indefinitely at a high performance of which human workers may struggle to achieve. That being the case, businesses that employ computers or machines over human workers can be more competitive in the marketplace and increase sales. JP Morgan utilised AI-powered software to go through 12,000 contracts in seconds, much faster than it would take a human, who require 360,000 hours to get through the same amount of contracts.

Furthermore, AI also allows computers to work with a kind of accuracy and precision potentially greater than any human possibly could. This can be particularly vital in medicine. Enlitic is a startup which shows this to be the case. The San Francisco-based company has managed to build a machine which examines images of x-rays and CT scans with the use of deep learning. By analysing masses of data, it can identify dangerous tumours within such images much more accurately than a human doctor. In a test against human radiologists, it, in fact, proved to be 50% more accurate with a false negative rate zero, compared to the 7% rate of the radiologists.

So, as well as very routine-based jobs being under threat by computerisation, AI is also causing significant shifts in more high-skilled work like medicine. Though, this is not just due to machines out-performing their human equivalents, the rise of AI is also having impacts in industries which have typically resisted the influence of technology. The legal profession is an example of this.

The recent advancements in AI allow computers to complete much more complex tasks in a much wider range of industries and professions unlike ever before. A range of different digital tools have emerged over the years taking over some of the tasks involved in legal work. E-discovery software, which is powered by AI, can go through heaps of legal materials for document review faster and more efficiently than a legal clerk or paralegal. Lex Machine supplies compilations of legal data, sourced from court cases and other legal materials, in a simplified and more easily digestible format (refer to the article titled 'Lawyers, You Are Not Immune' in the complimentary reading for more analysis on legal tech).

The disruption that could take place within the legal profession once again points to the cost-efficiency of AI. Legal fees have risen to a tremendous amount since the 1980s; in 2015, according to the Centre for Policy Studies, fees reached as high as £850 an hour. But with the emergence of AI-powered technologies, legal tech startups are able to provide much cheaper legal services in comparison to those offered by solicitors or other legal professionals.

With all that AI has to offer to the economy then, and the fact that its impact is much wider than previously thought, what chance do humans have in the workplace? Will technology eventually eradicate the need for human work altogether? It's these kinds of questions which reveal the divide on what AI's true impact on the future economy will be. In a paper by the Pew Research Centre, 65% of adults believe that computers will take over much of the work currently done by humans in the next 50 years. When asking experts on the subject, however, nearly half (48%) believe that "robots and digital agents will displace significant numbers" of high-skilled workers, whereas others (52%) disagree with this prediction. Those experts that disagree, though, do emphasise that while many jobs may indeed be taken over by robots or machines, such a movement may be compensated for by the creation of new jobs.

In the past, while improvements in technology have rendered some workers redundant, alternative work has managed to present itself. For example, the emergence of ATMs (automatic teller machines) provoked fears that the need for bank workers would become nugatory. Instead, ATMs helped to reduce costs for banks and consequently were able to open more branches in response to growing consumer demand which in turn required the recruitment of more staff. So, it is not necessarily the case that technology will eradicate jobs altogether. It could be the case that the introduction of new advanced technologies simply changes the nature of the work or even present alternative roles in the same industry.

With the rise of AI, there will certainly be a growing need for people with skills in coding the computer science. Additionally, these jobs could come with lucrative rewards; even engineers fresh out of university can earn around $120,000 a year. Lots of people in recent times have been recruited by the likes of Apple, Facebook and Alphabet to aid their respective AI projects, including self-driving cars and personal assistant apps. Thus, the alternative opportunities which may present themselves in the midst of AI's rise could be plentiful and bountiful.

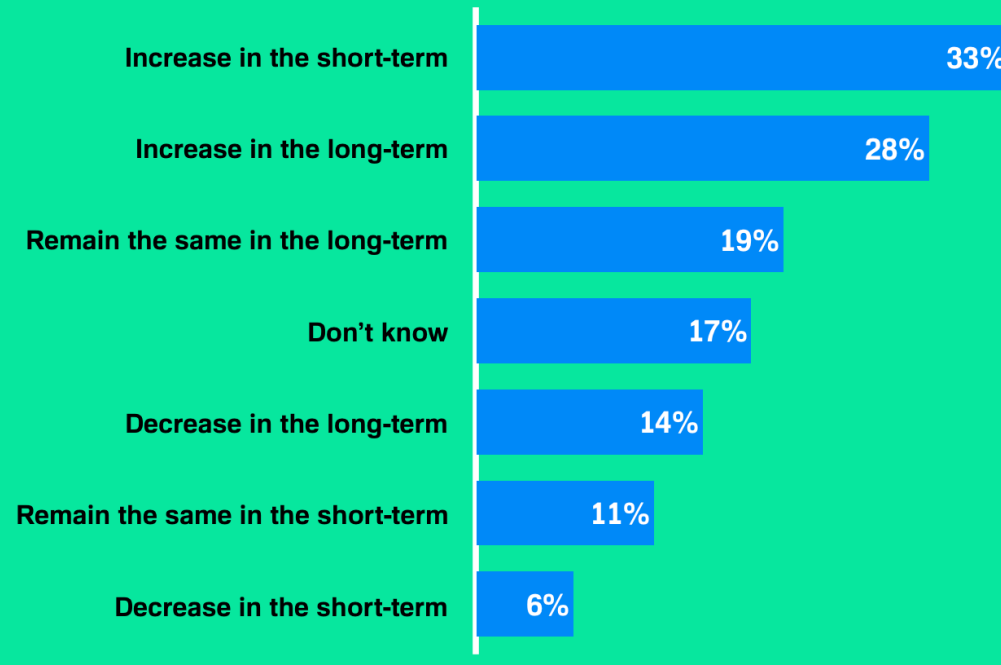
The flip-side to this, of course, is that becoming eligible for these roles may be difficult. The mass structural unemployment which could occur, if the impact of AI on the job market is potent enough, may be such that for those who lack the appropriate qualifications to transfer to computer-related work, alternative job opportunities may become hard to come by. This is especially so when such technology-related jobs are quite complex in nature

Although, even this could be seen as just a short-term effect. Over time, it is plausible to suggest that once economies adjust to the impacts of AI on the job markets, people will equip themselves with the right skills and qualifications to take advantages of the new jobs which may come into being. The younger generation today may thus, be in a better position to identify and prepare, while still in education, for the changes that improvements in technology may bring to their future world of work.

Overall, attempting to predict AI's true impact on the economy is a formidable task. Even so, to suggest that AI will eventually cause absolute unemployment, where nobody works at all, is an extreme allegation, to say the least. But even the more sensible predictions could be as equally probable since AI is an unprecedented technological phenomenon which may bring about unprecedented change. So, even relying on closely related historical precedents to provide the answer does not provide any great certainty. What can be said, however, is that AI will most likely eradicate some jobs, but not all jobs. This is what can be predicted for, at least, the foreseeable future, and any prediction beyond this will involve a great deal of speculation.

2

## How corporate security pros think AI will affect risk, 2015




Source: ISACA/RSA Conference State of Cybersecurity 2016

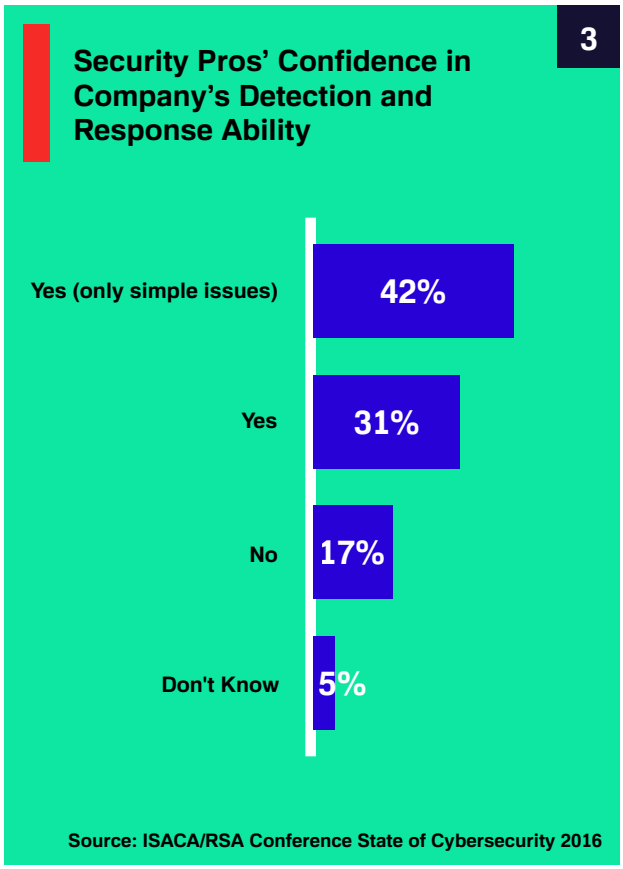
**Smarter Security**

If there is one industry which may benefit immensely from the implications of AI technology, it would be cybersecurity. With the power and capabilities that improved machine learning and algorithms could offer, potentially two major problems could be resolved to establish a safer and more secure cyberspace. The first concerns the main thrust of many cyber-attacks, which rely on manipulating the loopholes and weaknesses of human behaviour and habits, allowing hackers to deploy malicious software, steal precious data and corrupt networks. Secondly, the difficulty of building effective cyber defences, particularly with the constant evolvement and rapidly changing methods and schemes utilised by hackers, can be greatly reduced with the help of AI. As soon as one kind of attack is deflected, a new variation of it emerges, meaning that the development of security parameters is consistently out-paced and thus, out-mashed by the development of malicious software and the cyber criminals using them.

But one of the advantages of AI is that it is not subject to those natural habits and misjudgements that feature among humans. Due to this, AI-powered cybersecurity can provide better detection systems than any other human-operated systems. Startups such as CyberX and Dojo-Labs have managed to build programmes which can detect and block suspicious activity even on devices with low computing power. This is particularly needed with the explosion of IOT ('Internet of Things') devices in the past few years, many of which tend to lack even the most basic of security parameters, drastically widening the attack surface. The advent of AI-based detection systems will be able to analyse masses of data to improve accuracy and better differentiate between false alarms and real threats.

Cyber-attacks often come thick and fast, meaning that much smaller or less noticeable attacks can essentially go 'under the radar' while only the seemingly more obvious attacks are detected and dealt with. A report in 2014 by Damballa, a security firm, showed that there were approximately 10,000 security events taking place every day on the devices used by businesses. This number will most likely rise which would make the number of security alerts which pop up every day completely overwhelming for human operators. But one of the benefits of AI is its ability to analyse and learn from masses of data and information. CyberLytic is a startup using the power of AI to build detection systems for businesses to cope with these constant and growing threats. Its software uses machine learning techniques to classify attack data and successfully detect and deter threats. Such software has even been used by the UK government for its cyber operations. It is not so inconceivable to think that the need for a security team to carry out such tasks will become less critical as this technology improves.

One company which is taking advantage of implementing AI into cybersecurity is London-based Monzo, a smartphone based bank. This financial startup uses machine learning and big data to crackdown on financial crime and says that such techniques allow the company to stay ahead of the cybercriminals. Its fraud detection model, for instance, is based on Google's Tensorflow library (an




open source software library for numerical computation using data flow graphs) analysing numerous metrics, such as behavioural patterns, to identify potential fraudsters.

So far, it has been working. In June 2016, the company processed £5 million pounds worth of top ups and lost £316,000 to fraud, while in December 2016, it processed £21 million and lost just £40,000 to fraud. The digital bank has also managed to get better at distinguishing between genuine and fraudulent users; it now manages to ban only 1 genuine user for every 3 fraudsters as compared to its previous efforts of 6 for every 3. The use of machine-based learning detection system is how companies like Monzo are able to implement more accurate and dynamic cybersecurity that was much harder to achieve before.

But more than that, the use of AI allows Monzo to become a more secure bank without putting the burden of security on customers. One part of the payment process when using their app requires going through the 3D secure mechanism which, when enabled, redirects the customer to the bank that issued their payment card for further authentication when suspicious activity may be detected. This further layer of security, however, can sometimes worsen the user experience. Thus, Monzo has improved the accuracy of its fraud engine, not only meaning that customers have a much more smoother user experience, but also means that monthly financial losses are lower than 0.01% of the total top-up volume compared to 0.84%.

Aside from better detection systems, the use of AI can create cybersecurity that is more responsive and better at repairing vulnerabilities. Building security systems that not only more accurately detect breaches, but also use data about the coding structures of pieces of malware, or the exact features of the vulnerability in question, to produce patches and fill gaps in software, enable companies to cope with cyber-attacks far more effectively, especially zero-day attacks. Such systems would make companies far more responsive and adaptable to the improving sophistication of cybercriminals and hackers.

As beneficial as AI could be for cybersecurity, it will not eliminate all the problems and perhaps may not make the work of humans in the industry completely redundant, at least for now. Programmers and experts in security are still needed to not only build these programs but to maintain and improve them over time. Building software that has the ability to self-repair and self-improve are still in the early stages, and thus, there remains a need for specialists to continue working on such systems. While AI has improved drastically over the years, to build algorithms which can be left unsupervised, or at least partly supervised, has not yet been quite achieved, and there is perhaps still much of the journey left.

Furthermore, as many of these AI-powered security systems and parameters rely on masses of data, there accuracy and capabilities can only improve when this pool of data increases. The more data that can be analysed, more insights and patterns can be recognised to make these systems more accurate and more consumer friendly, much like the ambitions at Monzo.

Even so, could AI reach a point where machines understand humans more than humans themselves, which would, therefore, mean that the role of humans is unimportant? Cybersecurity is based on the behaviour of human beings in cyberspace, and in particular, how to protect against the more malicious behaviours and establish a sense of order and safety online which was never really considered at all in the first place. Up to now, humans were the only way to achieve a more secure cyberspace, since it was always accepted that humans knew best about humans. But AI is being combined with cybersecurity in a way that solves the problems and issues of the internet that humans have so far struggling to solve.

Cyberspace continues to grow, with IOT being playing a big part and thus widening the attack surface for hackers and other malicious online actors. The growth is at such a pace that humans have struggled to keep up, and ironically the work of humans has created more problems for the future. Therefore, in the context of cybersecurity, it is paramount that the necessary technologies are built and developed to cope with these dangers and evolutions, and AI seems to be providing those solutions.

**Know Thyself No More**

Interpreting the rise of AI as only having an impact on jobs, certain industries like cybersecurity or medicine or law is a limited view. If AI does, in fact, impact jobs and certain industries in the way that many have predicted, then its impact will inevitably be much wider. It will go beyond economics and cyberspace, as it will tap into the workings of greater society. If so, what does the future hold?

AI's influence on society's habits and lifestyles, as well traditional economic structures, has the potential to be tremendous. This matters because this will ultimately go on to affect our legal systems and what it means to be human with the introduction of another entity that mimics the

intelligence of humans. Before exploring these legal and policy implications, it is important first to look at what will be the foundation of these changes, which will be the societal impacts of AI as well as the other technological phenomenon that come with it, including Big Data and algorithms. There is even evidence today to suggest that the impacts of these technologies are indeed happening today, setting up a future where the role of humans could be very different. It will most likely challenge some of the key concepts and principles that have underpinned human existence for centuries.

The innovators of Silicon Valley and other tech hubs around the world are the ones responsible for pushing this agenda, albeit unintentionally perhaps. Nevertheless, the products and services they improve and release every now again are gradually doing more than disrupt businesses and industries. They are also disrupting people's lifestyles.

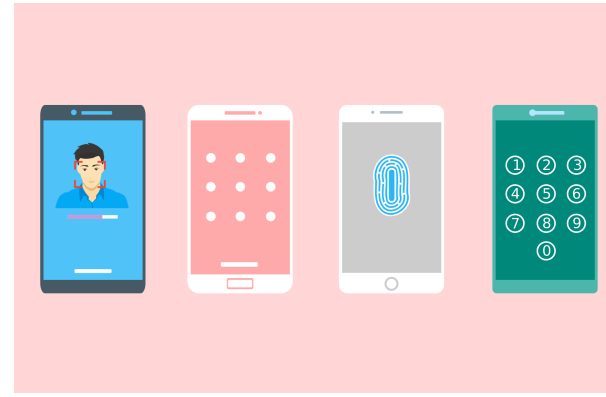
One idea in relation to this considers the impact of Big Data. The trend which has taken place in recent years has seen plenty of people heavily engrossed in the big network of data and information which flows around the world. There are two significant impacts this will have. The first concern is the traditional ways of thinking about privacy and how the embracement of Big Data will change this. The second is to do with the key ideas of individualism and free will; principles which humanity has stood by for a long while, but are now being challenged by advancements in technology.

Today's idea of privacy was quite accurately summed up by the late Steve Jobs; "Privacy means knowing what you're signing up for." This definition reflects the attitude of many consumers and business structures that play a role in the big flow of data that essentially runs the world. For many people, there is, of course, a line that is drawn between information one may want to keep private and some that one may want to release publicly. Increasingly, many of the apps and digital services being built today are shaped around this modern-day notion. Instead of requiring consumers to pay for services upfront, many companies have shifted towards a model where they can offer their services for free. But not entirely for free. Instead of payment in money terms, companies will now accept your personal data as payment. Facebook is a good example of this. You do not have to pay to use Facebook. Thus, the real value of a service like Facebook is the user's data. This is what is collected, analysed and used to sell ads on the site, which is, of course, one of the main ways that Facebook generates its revenue.

As such, it is very true that today consumers are willing to provide information about themselves in exchange for digital goods and services. Consequently, the concept of privacy in the age of data is simply one of a filter, where people are now wanting to contribute their data to the big flow, as opposed to just keeping it to themselves.

Of course, there will remain data that people will remain confidential, such as medical records. But with the improvements in AI and algorithms, people will want even this kind of sensitive information to be within the big data flow since doing so will provide critical insights and helpful advice.

This leads on to the second significant impact of AI on society, which is the decimation of free will and



individualism. It has long been held that humans had the monopoly on authority. In elections, politicians seek to appease to the wants of the people. In business, the customer is always right. Whichever party gets voted into office, or whichever product sells the most, are deemed correct or legitimate. This is because individualism supports the idea one should search their inner feelings to make decisions in their lives. The ingrained authority and value that can be found in one's inner-self is then transferred over to whatever decision they may make. Ultimately, human beings should believe in themselves. However, with the advances in AI, Big Data and algorithms, this idea is beginning to disintegrate.

Now that the world is starting to embrace machines and computers which are beyond the intelligence level of humans in certain respects, trust will be embedded into them instead of our inner feelings. This is because, in the end, our inner feelings are simply unquantifiable and subjective emotions which are often not entirely consistent. Why would humans revert to this when they can refer to machines that essentially know them better than they know themselves. When it comes to voting in an election, Facebook will gather up all your likes, comments and posts over time and analyse them to tell you which political party or candidate whose policies will best suit your preferences. On top of this, even the quantifiable tasks, such as seeing whether you have run out of milk, will be left to IOT devices and algorithms. If you do run out of milk, then your internet connected fridge will add it to your shopping order arranged to arrive at your house, later on, delivered by a self-driving vehicle that will also know what time you will be home to take the delivery. This is the kind of world that may very likely come into being, where the judgements of human beings becomes irrelevant as machines are entrusted to do the many tasks and chores we have become so accustomed to doing (refer to the article titled 'In Data We Trust' in the complimentary reading for more on this).

In his latest book called *Homo Deus: A Brief History of Tomorrow*, Yuval Noah Harari calls this trend the "Dataist revolution." Mr. Harari presents this increasing reliance on data and digital services like Google and Facebook as part of liberalism's demise:

*Naturally, Google will not always get it right. After all, these are all just probabilities. But if Google makes enough good decisions, people will grant it increasing authority. As time goes by, the databases will grow, the statistics will become more accurate, the algorithms will improve, and the decisions will be even better. The system will never know me perfectly, and will*

*never be infallible. But there is no need for that. Liberalism will collapse on the day the system knows me better than I know myself. Which is less difficult than it may sound, given that most people don't really know themselves well.*

One reaction to this may be to argue that this trend is over-hyped and highly exaggerated. It is fair to say that this data revolution may not happen, but there has been evidence to suggest otherwise. Amazon's flagship IOT product, the Echo, sold nine times more in December 2016 than it did the year before. The Financial Conduct Authority, a regulator in the UK, is concerned that big data may make insurance prices too expensive for clients, as insurers collect more data from social media and use better algorithms to analyse them and determine risk more accurately. This evidence is not exactly conclusive, but it does help to convey how society is gradually shifting towards a greater dependence on and trust in data and algorithms to provide meaning and the answers to life's complexities, big and small. If this is indeed the case, then the knock-on impact on law and policy will be tremendous too.

**Controlling the Creators and their Inventions**

Many technological inventions and innovations throughout human history are ethically neutral and apolitical. They are created and developed by human who they themselves may have political tendencies or agendas, but the technologies created are not necessarily themselves tailored to suit a specific agenda. This neutrality means that the inevitable political and ethical implications of these technologies will be dependent on how its creators use them.

The same can be said with the emergence of AI but to a certain extent. The unique feature of this technology is its potential ability to be independent of human control and act autonomously. In order to be able to do this, however, these machines and algorithms still depend on the activity and input of human beings to actually operate for the purpose that the AI entity has been programmed to perform.

Essentially, what can certainly be said is that the rise of AI triggers some unprecedented, and thus difficult, political debates, law and policy implications and ethical dilemmas. While the full effect is yet to be realised, there are a few impacts which can be identified today.

The first, which was alluded to in an earlier section, relates the implications on privacy and surveillance. In early 2017, Evernote, the maker of its popular note-taking app recognisable by its famous elephant logo, managed to highlight the difficulties that exist around the implementation of AI. In an effort to include machine-learning features into its digital services, it faced vicious backlashes from its users. The company had amended its privacy policy in order to pursue this implementation, making it clear to users in a blog post that Evernote employees would have access to information uploaded to its servers in order to operate the AI-powered features. Despite attempting to impress users of the new additions, and trying to reassure users by explaining employees would be "subject to background checks and receive specific security and privacy training at least annually," and that the data will be anonymised, Evernote failed to impress. After its announcement, users rebelled and demanded that the company reverse its decision citing an invasion of privacy. Subsequent to this disapproval, Evernote promptly changed its mind and apologised to users, promising to do better to adequately protect data and respect user privacy.

Although this happening may somewhat contradict the idea presented before in relation to users willing to give up some privacy in return for the convenience of technological innovation, this aforementioned user attitude is still in its relatively early stages. For now, may people will have the revelations by Edward Snowden of the mass surveillance conducted by intelligence agencies in America and the UK fresh in their minds. Therefore, privacy concerns, even in the context of the wonderful potential of AI and machine learning, remain alive (refer to the article titled 'The Elephant in the Room' in the complimentary reading for more on this story).

Consequently, such concerns may provide a barrier for businesses trying to embrace and implement AI into their products and services. In order for AI and machine learning to perform better, as much data as possible is needed. If users are apprehensive about offering their data to technology companies though, the potential of AI could be limited.

Aside from the private sector, the worries of the State using such technologies is perhaps of an even greater worry to some. For example, controversy arose in Russia with police using a programme called FindFace to identify individuals in public places. This programme collects data from social networks and machine learning to connect faces with online profiles, and thus determine their identity. Russian police have said that it has used the programme to track down criminal suspects or witnesses. The app's developers claim that such technology was intended to be used to identify people that one may see in a bar or on the street temporarily. Inevitably, privacy concerns have been raised, as the surveillance capacities of the Russian State are greatly enhanced by the introduction of such technology. This thus, shows how technologies developed by humans are truly impartial to political tendencies and ethical implications, and that these are dependant on how humans put the technology to use.

Another example involves the work being carried out by Google using AI. It is in the process of creating an AI-powered programme to filter out hate speech online. The software is called Perspective, and Google says it is an attempt to combat the vitriol trolling taking place online, and it can be used by websites owners to monitor the commenting systems. But as powerful and welcomed this technology may be, there is an ethical implication to be

realised. Google admits that the technology is in its early stages still, and so the programme has not proven to work too effectively yet. It is not yet managed to identify most of the hateful comments that exist online, while terms like "garbage truck" are identified as such. The ethical or policy considerations to be realised here is in determining what should be classed as hate speech. Of course, there is a wide consensus that racism, sexism and other kinds of discrimination are not to be tolerated. But how far should the programme go? Should a list of supposedly hateful phrases be decided just on Google's terms? What if the programmes conflict with First Amendment rights in America? And assuming that the software will be available globally as the internet would allow, how will it cope with the different laws and rules on hate speech in other jurisdictions? Europe, with its 'right-to-be-forgotten' laws, has a somewhat different approach to freedom of speech and expression than America, for its example.

In 2015, Google had a previous incident involving its Photo app which categorised pictures of black people as "gorillas." The software used AI to automatically group and label photos in the app. Richard Socher of MetaMind, an AI department now part of Salesforce, correctly pointed out that Google did not necessarily design the software to do this, but "if it trains on terrible data, it will make terrible predictions."

Such incidents highlight an interesting reality; the immense role technology companies are likely to play in the future due to the innovations they create. The more that their products and services interfere with the law and ethics, the more political tension they will infuse. The days where these firms could leave such problems to market forces are beginning to vanish. If there is anything to be learnt by the internet, with its lawlessness and borderless nature causing havoc for regulators, these tech firms will be held more accountable for their activity than ever before. AI will definitely be no exception.

**The AI Army**

One area which AI will certainly have a big impact, beyond the private and public sector context, is on international law and the rules of engagement. With all that AI can offer, it will inevitably be used in modern warfare, with similar advantages it can provide for cybersecurity being sought.

How exactly would the militarisation of AI be advantageous? There are a few ways it could be. Firstly, it can significantly reduce the need for boots on the ground. General Robert Cone, in charge of the US Army's Training and Doctrine Command, believes that drones and robots could reduce approximately a quarter of troops by 2030. Enhanced technology has allowed for an "a smaller, more lethal, deployable and agile force." As such, the US Army has contemplated cutting the size of brigade combat teams in the thousands. For now, many of the tasks which machines are capable of performing better than humans are specific and narrow. It is possible to envisage, however, that these machines will become smart enough to operate in a broader range of more complex scenarios and environments. Eventually, humans may not be needed to control and monitor these machines even from afar, and thus the likely consequence is that the need for human involvement in an increasing amount of conflicts will not be needed, at least

not to the same degree as today. As such, far less human lives have to be put directly in harm's way.

Secondly, AI entities may be capable of making much better decisions in the heat of the battle than any well-trained soldier could. As is with cybersecurity, AI entities, unlike humans, are not subject to emotions or irrationality, for they are just merely machines limited to only doing what they are programmed or instructed to do. Their natural 'calmness', if it could be phrased as such, means that the decision-making process is not subject to the occasional irrational behaviour or cognitive slips which humans are known to make, even amongst the most highly trained soldiers. The fact that AI entities are capable of processing much more data than humans also means that they are capable of making more informed decisions. In the same way that Alpha-Go was able to use Big Data to figure out the best moves to use in a game of Go, AI entities can use the same capabilities to determine the most effective operations and combat strategies to defeat the dedicated adversary.

This leads on to the third significant advantage, which is that AI entities would be able to better cope with the challenges presented by cyber warfare. This particular advantage is very reminiscent of the benefits of AI in cybersecurity in terms of its ability to deal with the growing amount of cyber-attacks taking place on a frequent basis and at an immense pace. Last year, the Defense Advanced Research Projects Agency (DARPA) at its Grand Cyber Challenge event tested autonomous hacking servers trying to attack each other while also patching themselves at the same time. Such abilities eliminate the problems currently presented by cyberspace, especially when it comes to adequately responding to attacks, which has become

immensely difficult due to their instantaneous nature. Leaving smart machines to defend as well as attack for themselves can mean a far more robust defence system can be put in place to attend to the growing modern threats.

Yet, in the context of cyber warfare, there are still some potential disadvantages to the rise of an AI army. In particular, there a few reasons to suggest that the militarisation of AI could pose even more severe problems than nuclear bombs did in the 1980s. The first identifies the structure of the internet and the impact it has on the traditional ways of conducting warfare. In cyber warfare, the ability to be one up on your adversaries is significantly harder than the other traditional domains (land, water and airspace). The internet's openness means that activity taking place on it can hardly be constrained effectively. This allows for a much wider range of adversaries. This is due to the fact that, whereas there may have only been a few countries capable of developing nuclear weapons since developing such weapons require costly raw materials which are hard to obtain, cyberspace allows information and resources to be much more ubiquitous and more easily accessible.

Thus, not only would there be more adversaries which military forces would have to cope with, but all of these adversaries can take advantage of the more even playing field created by cyberspace. The superiority of the America's cyber weapons and tools can more easily be outmatched by a hacker in Bulgaria. Therefore, as long as adversaries have a sufficient knowledge of code, a computer and a connection to the internet, they can, theoretically, be as capable of developing cyber offences as the US Cyber Command. The interconnectivity of the internet also means that an attack on one network can easily affect another, meaning the damage inflicted by an attack may not necessarily be limited to the intended target; the attack surface is thus much larger, expanding the dangers of cyber warfare. As such, the prospect of AI-powered electronic warfare could be far more damaging than nuclear weapons ever truly were, or at least as damaging.

A limitation to both the potential advantages and disadvantages of the militarisation of AI, though, would derive from the policy constraints on such weapons. The laws and regulations surrounding the development and use of autonomous weapons can perhaps necessarily frustrate the movement towards an AI army (it is necessary to note that the terms 'AI weapons' are 'autonomous weapons' are practically synonymous as they both essentially refer to a weapons system using information and data to independently select and engage targets). The applicability of the current legal framework surrounding autonomous weapons highlights two principles which may be of significant concern; distinction and proportionality.

Distinction is about the ability to distinguish between military and civilian targets. As said in the Intentional Court of Justice in 1996:

*The cardinal principles contained in the texts constituting the fabric of humanitarian law are the following. The first is aimed at the protection of the civilian population and civilian objects and establishes the distinction between combatants and non-combatants; States must never make civilians the object of attack and must consequently never use weapons that are incapable of distinguishing between civilian and military targets.*

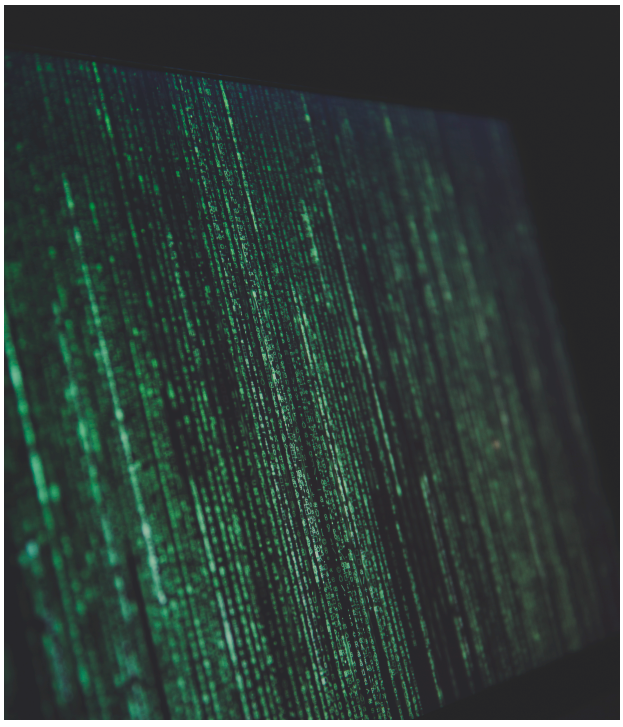
This important distinction principle is one of the important foundations of modern rules of engagement. The problem when it comes to autonomous weapons is whether they can be deployed and used in a way which adheres to such rules. Are these weapons capable of not carrying out indiscriminate attacks? As AI entities rely on masses of data to work accurately and effectively, the best way for autonomous weapons to adhere to the distinction rule would be to have access to quality data. This is where the difficulty lies. In her paper of autonomous weapons, Rebecca Crootof of Yale University presents the argument that autonomous weapons are not yet capable of distinguishing between civilian and military targets, as "doing so requires a complicated assessment of various factors, and there are many grey zones that bewilder even well-trained human soldiers." While civilians engaging in hostilities are lawful targets, "armed civilians acting as law enforcement" are not and distinguishing between the two is far from easy. Some predict that the ability of AI entities to differentiate between lawful and unlawful targets should improve in the near-future, but for now, the legal controversy in this regard remains problematic.

Similar legal incompatibilities are also observable with respect to the principle of proportionality. It is forbidden, according to the First Additional Protocol to the 1949 Geneva Conventions, to deploy an "attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated." When authorising an attack, determining whether it would adhere to the proportionality principle consists of subjective analysis. This is something which autonomous weapons, acting independently, may struggle to do. Crotoof correctly points out that autonomous weapons are not exactly "able to qualitatively analyse, let alone weigh, the expected military advantage of a particular attack and the associated potential harm to civilians." This is even more questionable when autonomous weapons are deployed in ever-changing environments, particularly where programming would have to be altered to adapt to such. The principle of proportionality is almost invariably centred around "human judgement," and so the objective nature of analysing data to determine whether an attack or a particular action is indeed proportionate is thus essentially incompatible with this legal principle. Accordingly, the use of autonomous weapons could be used in a way which contradicts some of the important rules and principles around conflict and war in a rather unprecedented fashion.

The military-oriented advantages of AI are promising in some ways while worrying in others. It may be difficult to say at this point which side it leans more towards, but it can certainly be said that it raises more questions than answers.

**Artificial Accountability**

A system of laws is designed to impose controls and boundaries on human behaviour and activity, in a way which promotes an orderly, equitable and prosperous society as much as possible. These boundaries, policed by judges and set by lawmakers, were always designed to apply to humans. But can these laws and rules be readily applicable to AI entities? If artificial intelligence allows



computers and machines to act and operate without human control, then the inevitable question arises as to who would be liable for their actions. Who exactly is accountable when an AI entity breaks the law or fails to comply with the rules?

The foundation of criminal law in many jurisdictions relies on two key principles which make up a criminal offence. The first is the *actus reus*, which is the conduct element of a crime. This conduct element can be achieved by an actual act or an omission, and such an act must be criminal to satisfy as part of the offence. For example, shooting a gun at someone, which brings about their death, would be the *actus reus* for murder. To complete the full offence, however, the *mens rea*, which is the mental element, has to be identified as well. If the person shooting the gun had the intention to kill the victim, then that person would fulfil both elements of the crime of murder. As it is necessary to find these two elements to show criminal liability, most of the time, how can this be applied to AI entities? In his paper, Gabriel Hallevy presents three possible models which may be used to impose liability on AI entities in criminal law.

The first model, of which Hallevy calls the 'Perpetration-by-Another' liability model, focuses on the concept of complicity. This model sees the AI entity as nothing more than a machine or a tool to commit a crime. The entity is ultimately seen as an innocent agent, and the capabilities of AI entities are thus understated. As Hallevy puts it, "[t]hese capabilities resemble the parallel capabilities of a mentally limited person, such as a child, or of a person who is mentally incompetent or who lacks a criminal state of mind." Under this model, the AI entity is treated the same as any other object which could be used to commit a crime. For instance, if a person uses a knife to stab someone else, the knife is not treated as an entity capable of criminal liability. Instead, the person using the knife is the one held criminally liable. That person would be liable as a perpetrator-via-another.

There are two limitations to this model. The first is who exactly would be the perpetrator-via-another. Would it be the programmer who created the AI entity in the first place, or is it the end-user who did not create the AI entity, but uses it for his own gain. To criminalise the programmer may seem a little remote; the programmer may have created the AI entity, but it may be hard to prove that he intended for that entity to commit a crime. Such an argument could be used to criminalise a mother for giving birth to a child who then goes on to commit an offence later on in life. Unless it can clearly be shown that the creator had an intention to create the entity for it to be used for criminal activity, this approach would seem harsh. Alternatively, it would appear more plausible in most cases to class the end-user as the perpetrator-via-another. This is because the end-user is the one who gives the instructions to the AI entity to pursue a course of action, which may be illegal, and thus, that user will be held liable for any criminal activity.

Yet, even this approach exposes another potential problem and also highlights the other limitation to this first model, which is that it does not acknowledge the advanced capabilities of the AI entity. As Hallevy points out, "[t]he Perpetration-by-Another liability model is not suitable when an AI entity decides to commit an offence based on its own accumulated experience or knowledge."

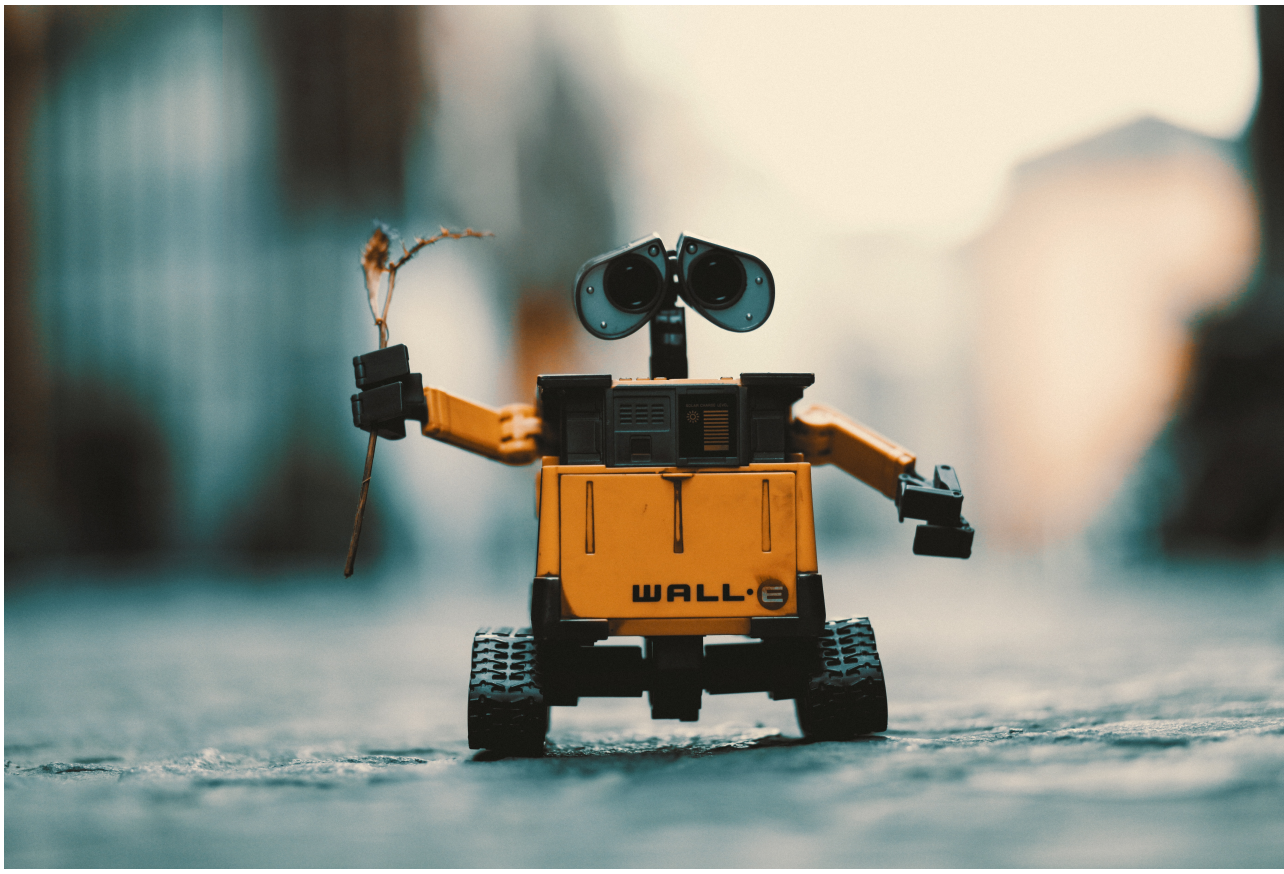
This leads on to the second model presented by Hallevy, called the "Natural-Probable-Consequence Liability" model. Hallevy defines this model as one which "assumes deep involvement of the programmers or users in the AI entity's daily activities, but without any intention of committing any offence via the AI entity." This can be demonstrated by conveying a scenario where a piece of AI-powered software is built to find and protect a computer system from threats on the internet. In the process of doing so, it discovers that it can find such threats by entering dangerous websites and destroying the threatening software. This would be a computer offence that the programmer, nor the end-user, did not necessarily intend for the AI entity to commit.

The natural-probable-consequence concept focuses on the idea of negligence. As such, it suggests that the programmer nor the end-user may have had the intention to commit the criminal offence, but that they knew that it might come about because such an offence is seen as a natural, probable consequence. It thus focuses on the implication that the negligence of either the programmer or the end-user is enough to be liable for the criminal offence.

Yet, the question arises as to whether the AI entity is still an innocent agent. If it is capable of acting independently, as suggested in the earlier example, then surely the AI entity no longer becomes an innocent agent. As such, there is an argument that as well as holding the programmer or end-user liable, the AI entity should also be held criminally liable as well.

This idea is dealt with in Hallevy's third model, called the "direct liability model." This model does not view the AI entity as an innocent agent dependent on a human programmer or end-user; the entity is observed as an independent entity thus capable of fulfilling both the conduct and mental elements of a crime on its own.

Under this model, the conduct element is the relatively uncontroversial part. As long as the AI entity takes a course of action which is part of the offence in question, and this




can be shown, then it can easily satisfy that element of the crime. This is also the case with an omission, where the entities failure to take action results in an offence. However, addressing the mental element of the crime, with regard to AI entities, can potentially be problematic. Are AI entities capable of the same thought processes as humans? Hallevy argues that they are. Humans use past experiences and consume information to take courses of action in the future. In the same way, AI systems use Big Data to enable the same processes, and so the two would not appear that different. As such, Hallevy argues that "the criminal liability of an AI entity according to the direct liability model is not different from the relevant criminal liability of a human," although he acknowledges that some adjustments should be made in certain cases. Hallevy also says that where the AI entity is affected by a computer virus or malware, or where it otherwise malfunctions, the entity should be able to rely on defences similar to those available for humans to mitigate its liability, such as that of insanity or intoxication.

If it can be established that AI entities can be held criminally liable as much as humans can, determining the appropriate punishment is also necessary. The most severe punishment for humans is the death penalty, as it completely eradicates the criminal for good, meaning that it becomes impossible for them to reoffend. Such a punishment is not so applicable to AI entities. Even if an AI entity is deleted, it is not gone indefinitely. As Hallevy correctly points out, "[t]he arrangement of the code that makes up the convicted AI entity may be gone momentarily, but another programmer can discover and imitate the same arrangement of code to bring the AI back to life essentially."

Similar problems arise with imprisonment. For human beings, such a punishment rids them of their liberty and freedom. But how exactly can an AI entity be locked up for the same effect? Is it at all the case that AI entities would feel their freedom and liberty compromised in a way which makes them feel like they are being punished? Hallevy even accepts that "humans have feelings that cannot be imitated by AI software, not even by the most advanced software." Thus, if the equivalent punishment for AI entities is to suspend them from their work, would this really achieve the effects that imprisonment of human beings is supposed to have. The effect of punishment is to impose hardship or rough treatment to essentially force the recipient to recognise that the offence they committed was wrong so that they will not reoffend. If AI entities are not capable of feeling subject to such hardship, it may be difficult to see how such any punishment could be truly effective.

If AI entities are to exist in the world we live in, it will make sense for them to adhere to the standards, rules and conventions put in place in order for them to operate harmoniously. Thought should thus be given to how the criminal liability of AI entities should be applied, as their continuing developments and improvements start to have an impact in unfathomed ways.

**Robot Rights?**

There is an argument to be made that AI entities, as they improve and start to replace humans in a number of capacities, should be subject to certain rules and regulations so that they continue to operate for the benefit of humanity. In that case, is there an argument to be made to suggest that

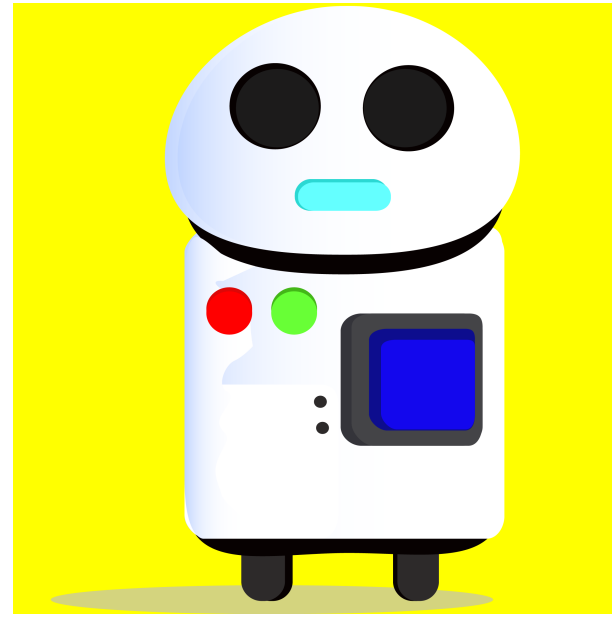
AI entities should be able to enjoy rights as well as responsibilities?

It may be helpful to analyse the reasoning behind giving human rights before suggesting that AI entities should enjoy the same rights or something equivalent. There is an understanding that the rationale behind human rights lies in the concepts of dignity and consciousness. Article 1 of the Universal Declaration of Human Rights, says "[a]ll members of the human family are born free and equal in dignity and rights…" There is thus, the implication that all humans are born with an inherent dignity, of which should not be overridden by others and which should be acknowledged and valued indefinitely.

Underlying this concept of dignity, though, is that of consciousness. The exact recognition or definition of this has never been made entirely clear, yet its mysterious existence does not prohibit its seeming importance to humanity and rights. The basic idea of consciousness is that humans are aware of themselves and their feelings. As such, humans have the ability to suffer. Pain and suffering are instinctively feelings which humans avoid as we are programmed to treat these feelings and emotions as potential threats or warning signs to our survival. This is why we know not to touch a hot fire. Consciousness also means that humans have the ability to be happy. The Founding Fathers of the United States sought to limit the power of the state to allow the people to engage in the pursuit of happiness, according to Yuval Noah Harari. Being aware of ourselves and our feelings means that we have the ability to seek out what we believe is best for us to extend our survival and secure our happiness. Thus, one of the major purposes of human rights is to safeguard our happiness, comfort, dignity and survival while diverting anything which may threaten this, such as torture.

As such, it is difficult to see how this construction and understanding of human rights can be at all applicable to robots or AI entities. Yuval Noah Harari even points out that "[r]obots and computers have no consciousness because, despite their myriad abilities, they feel nothing and crave nothing." AI entities cannot feel pain, depression, sadness or suffering, things which human rights are meant to curtail to a certain extent. Therefore, there seems to be little point in granting rights to AI entities or robots that are not conscious like humans are.

However, Dr. Kate Darling, a robot ethicist from MIT's Media Lab, said in an interview with PC Mag that the way that humans interact with technology and the choices we make in how we use our technology could be a reflection of ourselves. The argument here is that humans may have an inclination to treat other living things which are not human as though they are human. Although, this inclination is not necessarily due to the belief that those living things are conscious and thus, deserve rights, but that if they are not treated with dignity, then this reflects human beings as cruel and savage. This idea departs from the idea of 'human exceptionalism,' which purports the idea that humans come before anything else. Immanuel Kant, a German philosopher, even suggested that this seemingly inherent uniqueness is overstated and instead suggests that there may be other beings which are like us. As such, humans may not be as special as perhaps believed to be to then warrant any other living thing to be considered secondary.



Could AI entities be deserving of rights as humans are? It would appear so. In 2016, a report by the European Parliament recommended the creation of "electronic personhood" which would award rights and responsibilities to advanced AI entities. Yet, even if taking this direction with regard to robot rights is driven by a desire to convey Homo sapiens as an empathetic race, granting rights to AI entities still contravenes the principles underlying the concept of rights. Even the most advanced forms of AI are not capable of developing emotions as humans are. Does turning off an AI-powered machine harm the machine, or make it sad or feel abused? The consensus would likely be no, yet this contention is being challenged. In 2015, a Japanese telecom company called Aldebaran Robotics built a robot capable of feeling human emotions like joy and anger. The use of sensors, cameras and other forms of input provide the bot with data which it uses to react to certain scenarios using the appropriate emotions. The bot essentially creates its own emotions using an "endocrine-type multi-layer neural network," emulating the behaviour of humans.

But even if we could create robots which could feel and display human-like emotions, would it be desirable to do so? It is hard to imagine a scenario where humans would want to create AI entities or robots which were conscious. If an AI entity cannot feel hunger or tiredness, then that means it can work 24/7 without taking a break or sleeping, unlike humans. Consequently, from an economic perspective, the AI entity could be highly productive, to the benefit of the business and the consumer, as more goods are made and sold, more services are provided, and none of it has to ever stop, and no wages have to be paid. It creates an economy of almost costless production. It might be more plausible to say that humans will only build AI entities which benefit humankind itself, and not necessarily focus so much on the well-being of the AI entity. As Amitai and Oren Etzioni argue in their article, "[h]owever smart a technology may become, it is still a tool to serve human purposes." This consensus is what fuels the development of self-driving cars for example, as it is envisaged that such a technological advancement will reduce car accidents and traffic on the

roads, bettering human lives. The control humans have over the creation of AI entities means that the question of them becoming conscience is essentially a human decision. If developing bots with consciousness is not demanded, then it will not exist.

Yet, while humans may stay committed to developing AI entities solely for the good of humanity, and thus disregarding any rights an AI entity could have, is there any plausibility to the idea of AI entities eventually becoming smart enough to develop their own AI entities which would then demand such rights? Amitai and Oren Etzioni make the point that if AI-powered technologies become increasingly sophisticated, "[t]hey may…act in defiance of the guidelines the original programmers installed." If AI entities gradually drift away from the original creator's intentions, those entities may build further robots and entities and thus embark on a movement which allows the rights and interests of AI entities and robots to prevail over those of their human creators. But these predictions are more identifiable in sci-fi movies and tv shows and the probability of this prospect actually materialising is perhaps quite low, at least for now. Therefore, these propositions are certainly questionable but necessarily impossible.

Perhaps only in this way can robot rights be firmly established and be on par with human rights. Unless we are prepared to forgo our unique 'specialism' which has been central to our living and understanding of the world, then the idea of robots or AI entities obtaining rights seems far-fetched. As long as humans treat them as mere tools which are confined to the desires and intentions of human beings, then the argument in favour of robot rights remains weak, but it should not be completely disregarded.

**Where I Apply**

AI has proven to be a rapidly improving technology applicable to so many aspects of our lives. It has the potential to drastically change the economy, the armed forces, law and policy and other several other important parts of the world. As such, it begs the inevitable question; what will be humanity's role in the age of AI?

The simple answer is that it depends. Whether the development of AI and robotics becomes a force for good or bad ultimately depends on the choices that are made. The technicians, computer scientists and technology companies will clearly be major stakeholders in the midst of AI's rise. But lawmaker and world leaders will also need to have critical part to play, as they very often do with regard to technological innovations.

Yet, with that, what should be avoided is governments and lawmakers being too slow to respond to these rapid innovations, which has been a more familiar theme in recent times. Cyberspace has now crept into so many aspects of our lives without any vote on how it should work and how it should be regulated. The consequence is that the free market takes responsibility for these decisions and, as such, privacy and security concerns of the online world were not paid attention to early on, and today they have become some of the greater challenges of cyberspace today. That is not to say that government regulation is always better than free markets, but when the capitalist-driven innovative mindset of Silicon Valley based technology firms overrides the safety-conscious mindset which should also be exercised, not all the technological innovations created can be for the good of humanity in the long run. As Yuval Noah Harari argues, â€œ[i]t is dangerous to trust our future to market forces, because these forces do what is good for the market rather than what is good for humankind or for the world.

Therefore, the future of humanity ultimately lies in humanity's hands. If the consensus amongst those who continue to fuel the rise of AI is that humans will eventually become redundant in the many aspects of the world as we know it, then that may very well be the fate of humanity in the age of AI. If this possibility materialises, then one unfortunate consequence may be that the fate of humanity lies in the hands of a relatively small group of people, namely the technicians, computer scientists and technology companies producing these technologies. They could very well be perceived, in the same way that many view the banking industry today, as selfish elites whose interests do not focus on the people their activities impact. Instead, these elites, whose work essentially determines the fate of much of the world, care only about their own well-being and not the bigger picture.

As such, technology companies today are often under the spotlight in the aftermath of terrorist. They are criticised for not doing enough to crack down on terrorist-related content and the use of social media and the internet by militant groups to recruit and influence potential sympathisers abroad. Since the government and regulators did not impose their authority on the internet from the beginning, attempting to do so now in an effort to strengthen security becomes much more difficult, since a significant part of this task lies in the hands of those technology companies.

The rise of AI, thus, should be accepted with caution, but not perceived as a guaranteed threat to humanity. In June 2017, it was reported that Google was developing AI-powered software to block terror posts. Thus, in order for AI to be a force for good, the right decisions have to be made. In order for this to happen, it would be ideal to have both lawmakers and technologists to make this a reality, working together to find reasonable solutions to the problems ahead. But if the rise of AI follows the same path as the internet, the effects could be detrimental. ■

Complementary reading:

- Lawyers, You Are Not Immune
- The Elephant in the Room
- In Data We Trust!
- When Politics Bytes

These articles can be found online at thecybersolicitor.com

# Bibliography


**Introduction**

1. Economist.com. (2016). Showdown. [online] Available at: http://www.economist.com/news/science-and-technology/21694540-win-or-lose-best-five-battle-contest-another-milestone
2. What's the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning?. (2016). [Blog] Nvidia. Available at: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/
3. Ft.com. (2016). Yuval Noah Harari on big data, Google and the end of free will. [online] Available at: https://www.ft.com/content/50bb4830-6a4c-11e6-ae5b-a7cc5dd5a28c

**Bigger, Better and Faster**

1. Science, L. (2017). A Brief History of Artificial Intelligence. [online] Live Science. Available at: http://www.livescience.com/49007-history-of-artificial-intelligence.html
2. Wired.com. (2014). The Three Breakthroughs That Have Finally Unleashed AI on the World. [online] Available at: https://www.wired.com/2014/10/future-of-artificial-intelligence/
3. En.wikipedia.org. (n.d.). Artificial intelligence. [online] Available at: https://en.wikipedia.org/wiki/Artificial_intelligence.
4. Nielsen, M. (2015). Neural Networks and Deep Learning. [online] Neuralnetworksanddeeplearning.com. Available at: http://neuralnetworksanddeeplearning.com/chap1.html
5. Harvard Business Review. (2012). Big Data: The Management Revolution. [online] Available at: https://hbr.org/2012/10/big-data-the-management-revolution

**Computer Capital**

1. Economist.com. (2016). Automation and Anxiety. [online] Available at: http://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety
2. AI, Robotics, and the Future of Jobs. (2014). 1st ed. [ebook] Pew Research Center. Available at: http://www.pewinternet.org/files/2014/08/Future-of-AI-Robotics-and-Jobs.pdf
3. Economist.com. (2017). Will robots displace humans as motorised vehicles ousted horses?. [online] Available at: http://www.economist.com/news/business-and-finance/21719761-probably-not-humans-have-lot-learn-equine-experience-will-robots?frsc=dg%7Cc.
4. Lawyers, You Are Not Immune. (2016). [Blog] The Cyber Solicitor. Available at: https://www.thecybersolicitor.com/single-post/2016/08/19/Lawyers-You-Are-Not-Immune
5. Economist.com. (2016). Tech firms shell out to hire and hoard talent. [online] Available at: http://www.economist.com/news/business/21709574-tech-firms-battle-hire-and-hoard-talented-employees-huge-pay-packages-silicon-valley
6. Susskind, R. and Susskind, D. (2017). The Future of the Professions. 1st ed. Oxford University Press
7. Economist.com. (2017). Machine-learning promises to shake up large swathes of finance. [online] Available at: http://www.economist.com/news/finance-and-economics/21722685-fields-trading-credit-assessment-fraud-prevention-machine-learning.

**Smarter Security**

1. Yakir Golan, M. (2016). How AI will transform cybersecurity. [online] VentureBeat. Available at: https://venturebeat.com/2016/11/22/how-ai-will-transform-cybersecurity/.
2. Gent, E. (2016). Battle of the Bots: How AI Is Taking Over the World of Cybersecurity. [online] Singularity Hub. Available at: https://singularityhub.com/2016/11/09/battle-of-the-bots-how-ai-is-taking-over-the-world-of-cybersecurity/.
3. Darktrace.com. (2016). Darktrace | Darktrace Cyber 'Immune System' Fights Back. [online] Available at: https://www.darktrace.com/press/2016/84/.
4. Monzo.com. (2017). Fighting Fraud with Machine Learning. [online] Available at: https://monzo.com/blog/2017/02/03/fighting-fraud-with-machine-learning/.
5. Millman, R., Reeve, T. and Winder, D. (2016). Artificial intelligence and the future of cyber-security. [online] SC Media UK. Available at: https://www.scmagazineuk.com/artificial-intelligence-and-the-future-of-cyber-security/article/531621/.
6. University, L. (2016). Artificial Intelligence Toolkit Spots New Child Sexual Abuse Media Online. [online] Lancaster.ac.uk. Available at: http://www.lancaster.ac.uk/news/articles/2016/artificial-intelligence-toolkit-spots-new-child-sexual-abuse-media-online/.
7. Anon, (n.d.). TensorFlow. [online] Available at: https://www.tensorflow.org.

**Know Thyself No More**

1. thecybersolicitor.com. (2017). In Data We Trust!. [online] Available at: https://www.thecybersolicitor.com/single-post/2017/03/17/In-Data-We-Trust.
2. Harari, Y. (2017). Homo deus. 1st ed. [Place of publication not identified]: Vintage, Location 5146 (Kindle).

**Controlling the Creators and their Inventions**

1. Ft.com. (2017). Frankenstein fears hang over AI. [online] Available at: https://www.ft.com/content/8e228692-f251-11e6-8758-6876151821a6.
2. Economist.com. (2016). Frankenstein's paperclips. [online] Available at: http://www.economist.com/news/special-report/21700762-techies-do-not-believe-artificial-intelligence-will-run-out-control-there-are.
3. Gershgorn, D. (2017). Alphabet's hate-fighting AI doesn't understand hate yet. [online] Quartz. Available at: https://qz.com/918640/alphabets-hate-fighting-ai-doesnt-understand-hate-yet/.

4. thecybersolicitor.com. (2017). The Cyber Solicitor. [online] Available at: https://www.thecybersolicitor.com/single-post/2017/01/30/The-Elephant-in-the-Room.
5. Ft.com. (2016). Algorithms are not impartial — and may penalise by race. [online] Available at: https://www.ft.com/content/c90e68a4-661d-11e6-8310-ecf0bddad227.

**An AI Army?**

1. Crootof, R. (2015). The Killer Robots Are Here: Legal and Policy Implications, 1–79.
2. REVIEW, T. C. D. (2016). The Inevitable Militarization of Artificial Intelligence, 1–13
3. Rejcek, P. (2016). Killer Robots Won't Go to War If Global Movement Has Its Way. [online] Singularity Hub. Available at: https://singularityhub.com/2016/12/11/killer-robots-wont-go-to-war-if-global-movement-has-its-way/
4. Military, F. (2014). Are robots replacing human soldiers?. [online] HowStuffWorks. Available at: http://science.howstuffworks.com/robots-replacing-soldiers.htm.
5. Work, R. O., & Brimley, S. (2014). 20yy Preparing for War in the Robotic Age, 1–44.
6. Dinstein, Y. (2012). The Principle of Distinction and Cyber War in International Armed Conflicts. Journal of Conflict and Security Law, 17(2), 261–277

**Artificial Accountability**

1. Hallevy, G. (2010). The Criminal Liability of Artificial Intelligence Entities, 1–42.

**Robot Rights?**

1. Kurzgesagt - In a Nutshell. (2017). Do Robots Deserve Rights? What if Machines Become Conscious?. [online] Available at: https://www.youtube.com/watch?v=DHyUYg8X31c.
2. Harari, Y. (2017). Homo deus. 1st ed. [Place of publication not identified]: Vintage, Location 1632 (Kindle).
3. Etzioni, A., & Etzioni, O. (2016). Designing AI systems that obey our laws and values. Communications of the ACM, 59(9), 29–31. http://doi.org/10.1145/2955091
4. Medium. (2017). Do Robots and AI Deserve Rights? – PC Magazine – Medium. [online] Available at: https://medium.com/pcmag-access/do-robots-and-ai-deserve-rights-1ae1fbf89f69.
5. Perry, Michael J., The Morality of Human Rights: A Nonreligious Ground?. Emory Law Journal, Vol. 54, pp. 97-150, 2005.
6. Louden, R. (2014). *Kant's human being*. 1st ed. Oxford [u.a.]: Oxford Univ. Press, p.20.
7. Murphy, M. (2015). *Robots in Japan now have emotions*. [online] Quartz. Available at: https://qz.com/433877/robots-in-japan-now-have-emotions/

**Where I Apply**

1. Harari, Y. (2017). Homo deus. 1st ed. [Place of publication not identified]: Vintage, Location 5701 (Kindle).
2. Ft.com. (2017). Facebook turns to AI to help block terror posts. [online] Available at: https://www.ft.com/content/93f2f9b2-51d8-11e7-bfb8-997009366969.

# THE CYBER SOLICITOR

Vol No. 1

Issue No. 2

June 2017