Scientists for Global Responsibility

Scientists · Architects · Engineers · Technologists

Contents

Artificial intelligence: how little has to go wrong?

Autonomous weapons, driverless cars and friendly spies in the home

A Responsible Science briefing & survey

Published by Scientists for Global Responsibility (SGR), part of the campaign awarded the Nobel Peace Prize 2017.

Written by Andrew Simms with Dr Stuart Parkinson

Andrew Simms is assistant director of Scientists for Global Responsibility, co-director of the New Weather Institute, a research fellow at the Centre for Global Political Economy, University of Sussex, and a fellow and the former policy director of the New Economics Foundation. He is the author of several books including, *Cancel the Apocalypse: the new path to prosperity.*

Dr Stuart Parkinson is executive director of Scientists for Global Responsibility. He holds a PhD in climate science and has researched and written widely on a range of ethical issues related to science and technology, including climate change and military technology. He is lead author of the SGR report, *Offensive Insecurity*.

Published by Scientists for Global Responsibility (SGR) in November 2018.

Printed copies of this report can be obtained from:

Scientists for Global Responsibility,
Unit 2.8, Halton Mill, Mill Lane, Halton, Lancaster LA2 6ND.

Email: info@sgr.org.uk

Electronic copies can be downloaded from: http://www.sgr.org.uk/

Introduction: AI – on tap or on top?

Is the human handling of artificial intelligence smart? Is it set to solve many of our problems or open a Pandora's box? Increasingly complex algorithms linked to automation have potential applications in almost every imaginable sector. Autonomous mechanisation and machine learning are being insinuated already, often barely noticed, into the world around us.

But is the science on tap or on top? Are we in control, or playing a kind of technological Russian roulette in which we spin the chamber of autonomous learning and / or decision making by machines without being fully in control of what happens when the trigger is pulled? The weaponised metaphor is not hyperbole, because one of the most controversial issues surrounding AI is in the field of warfare and military technology.

Applications are, however, emerging and being introduced in places ranging from on the road with driverless cars, to in the home and workplace with digital assistants, and from farm to hospital. There are so many issues that this briefing aims only to describe some of their breadth.

Potential applications of AI in healthcare, for example, have been widely reported.[1] The ability to spot patterns and aid diagnoses seems promising. But in this briefing, we aim to raise questions on some potential implications which are more worrying or have attracted less attention than others. To do that we have also surveyed members and supporters of Scientists for Global Responsibility (SGR), an organisation comprised mostly of scientists, engineers and technologists. The results suggest that, at the very least, far too little scrutiny and public debate has accompanied the creeping deployment of AI and a rethink is needed. SGR is not alone in believing this, especially where the issue of military applications is concerned, where a large, international body of opinion is calling for a moratorium on the development of autonomous weapons.

Less stark, but equally important in other ways, many feel that the encroachment of AI into aspects of our home and working lives, not to mention our roads, communication and operating systems, and even our countryside needs to be subject to much more responsibility and careful consideration.

Headline issues

> *"The biggest danger is that we use these tools to entrench our existing biases and compound the injustice that we already see in the world around us. These systems are trained on past human data, and that data represents the same kind of injustices that we would like to fix today. So what they are likely to do often, is to reproduce the biases and the blindspots that we as humans already had in running our existing systems, and that has been recorded in the data and subsequently used to train the algorithms."* [2]
>
> **Mustafa Suleyman, Co-founder and Head of Applied AI at DeepMind, 6 September 2018**

Weaponisation

One of the starkest manifestations of AI is the danger of an arms race in lethal autonomous weapons.[3] The debate in this field goes beyond, but is conditioned by, the morality and accountability of the use of drones in warfare, which itself harks back to an earlier debate on aerial bombing.[4] Alongside threats to privacy and jobs, this is one of the most feared potential developments of AI and autonomous machine technology. It has mobilised researchers such as those who signed a letter calling for an outright ban on autonomous weapons. The letter stated that, *'If developed, autonomous weapons will be the third revolution in warfare. They will permit war to be fought faster and at a scale greater than ever before. They have the potential to be weapons of terror. Despots and terrorists could use them against innocent populations, removing any ethical restraints. This Pandora's box will be hard to close if it is opened. As with other technologies banned in the past like blinding lasers, we can simply decide not to develop them.'* [5] Technology already exists, for example, potentially to create an autonomous lethal drone linked to facial recognition technology, which could act as an assassination tool. And, car bombs could be made to self-detonate when concentrations of people close-by were highest.

The illusion of quick wins

In 2014 the Pentagon announced what it called its 'Third Offset Strategy.' By using big data (very large data sets that when analysed, reveal patterns, trends and

associations), autonomous systems and robotics, it would retain its global military dominance. This promised the 'clean', 'smart' and 'precise' use of force. But we've been here before in terms of military technology creating the illusion of superiority and promising quick wins.

As *The Economist* recently wrote of military thinking prior to the World War I, rapid deployment of troops and equipment by rail coupled with new, quick firing artillery was meant to make all wars henceforth rapid and decisive. Of course, this is not what happened. Next, when air forces began delivering heavy bombardment from above, it was thought this would lead to rapid surrender, but the World War II bombings of cities like London, Dresden and Tokyo proved that idea wrong as well. The scale of destruction did however increase considerably along with civilian casualties.[6] More recently, Western nations' overconfidence in their technological superiority in the wars in Iraq and Afghanistan have contributed to prolonged conflict in those countries as well.

Historian and former US Marine, Prof. James Simms writes of the WWI case that the flawed thinking was partly because military strategists predicted the consequences of the latest battle technology based on analysis of how previous wars had been fought. The consequences were that 'newer, more lethal weapons had industrialized the practice of war, expanding the relatively compact battlefields of earlier ages into vast killing fields.'[7] Military and political leaders made the assumption that new technologies would shorten conflicts, not lengthen them, as happened through the use of entrenched defensive positions. It was thought too that the economic cost of war would keep them short, but instead combatant nations converted domestically to war economies – and, to some extent, boomed. Indeed, large military corporations have a long history of exploiting war for financial gain – and retain a large degree of influence over governments even in some democratic nations.

Today, in terms of the containment of conflict, the potential of autonomous weaponry, such as armed drones, leaves borders more porous than ever, and the potential scale and reach of conflict without restraint.

As with other forms of arms, a global trade in small, cheap, armed drones could be driven by despotic government use as well as non-state organisations. Friendly drones could be hacked leading to unpredictable effects. Drones could be programmed to target particular groups and would do so, but with high error rates. One reason why the US supported the chemical weapons treaty – having used such weapons in Vietnam (for example, Agent Orange) – was because they came to accept that they are a 'poor man's weapon of mass destruction', and therefore more useful to enemies than to the US.[8]

If widely deployed by states, autonomous weaponry would inevitably be taken up by non-state actors contributing to a greater, generalised terror. This, coupled with high risks of proliferation, explains widespread calls for an outright ban.

Accidental nuclear war

Some of the greatest risks lie where the greatest power is concentrated. This is literally the case where nuclear arsenals and their control systems are concerned.[9] The increasing use of 'Computer Network Operations' by militaries to try to gain an advantage by, for example, infiltrating and disrupting an 'enemy's' nuclear weapons systems could be enhanced by, and the risks increased by, the use of AI. Command and control systems for nuclear weapons are becoming increasingly automated. Hundreds of US and Russian nuclear weapons systems are still kept at their Cold War 'high alert' status – technically known as 'launch on warning' - whereby a nuclear weapons attack can be ordered in response to the apparent detection of launch of the enemy's nuclear weapons.

The decision window for heads of state, such as a US or Russian president, is a matter of minutes. But the information available to make that decision can be corrupted by false sensor readings, computer glitches and human error. Add to this, deliberate attempts to corrupt the decision-making process by malware - which could become increasingly sophisticated if it incorporates more advanced machine learning abilities - and the risks are heightened.

It is very hard for citizens to know what secret military cyber capabilities are being developed or used, but it may also be equally difficult for elected representatives and regulators. It is possible that their existence may only become known if they are deployed. For example, the Stuxnet virus used in 2010 to infiltrate and cause major damage to isolated, underground Iranian nuclear facilities, was almost certainly developed and deployed by the US (and Israeli) military.[10] Up until that point, such attacks were thought to be in the realm of science fiction.

To insulate nuclear systems they are often air-gapped – physically isolated - from the wider internet and networks that are not secure, but the Stuxnet attack showed that ways can be found to "jump the air gap." Malware can be hidden inside physical components, and through upgrades and routine maintenance corrupted components can being fitted. While the electronics industry is global, it is very hard to secure this. In 2015 the US National Security Agency was revealed to have hidden 'spyware' in computer hard drives used in a range of countries, while in 2018 China was reported to have used tiny chips to hack American companies.[11]

And whatever one state might develop to use against another, it always places itself at risk of having something similar used against it.

Rogue intelligence

Data harvesting and fake news

It's long been understood that information is power. Sometimes the full implications of what data can do hasn't been realised until after it has been collected. The supermarket Tesco's stole a march on its competitors with its Clubcard loyalty scheme. Intelligence gathered on customers allowed them to 'intelligently' target customers with marketing.[12] By knowing what, when, where and how frequently people were buying different things allowed them through triangulation to draw highly personal conclusions, such as whether or not a customer might have certain health or addiction problems, or whether they might be pregnant. If a customer also banked with the supermarket, and used its mobile telephone services too, an extraordinary detailed and intimate picture of that person's life would be in the hands of a large, commercial corporation. But debate concerning the insidious use of data and AI to influence human choices as both consumers and citizens was fully flushed into the mainstream by the scandal of Facebook's involvement with the controversial lobbing and public affairs firm Cambridge Analytica.[13]

Now insolvent as a result of revelations, the firm used data harvested from Facebook to inform the services it provided to political clients, with the objective of influencing the outcomes of elections. While Facebook said it had been misled by Cambridge Analytica, a UK parliamentary select committee found that Facebook had allowed its platform to be abused in a range of ways, including by the Russian government, to influence the outcome of US elections.[14] The use of 'bots' on social media to distribute 'fake news' has become a focus of concern around the dangerous and unethical use of the technology. Speaking around the launch of a report on the subject from the House of Lords, Lord Clement-Jones said, "Whether or not the data analytics they carried out was actually using AI … It gives an example of where it's important that we do have strong intelligibility of what the hell is going on with our data."[15]

State insecurity

The State has sought as comprehensive a picture of its citizens as possible in the UK at least since the compilation of the 11th century Doomsday Book in Norman times. But the ability to track individuals is today beyond previous imagining. Problems quickly arise when dependence on a massive convergence of gathered data combines with AI applications and machine learning.

For example, law enforcement and state security services are increasingly turning to facial recognition technology. But that has proved to be highly inaccurate in practice, creating two key problems. The first is that this undermines the effectiveness of the jobs those services are charged to perform. The second is an equally worrying risk of inadvertently putting innocent people under suspicion and wrongly implicating them in crime. This would also blemish their data record.

In one piece of research, facial recognition technology used by Britain's largest police force, the Metropolitan Police, failed spectacularly – producing 'false positive' identifications 98% of the time. In another instance, a system used by the South Wales Police produced 2,400 false positive results across 15 uses from June 2017 onwards. Its success rate was lower than 10%.[16] The government's information commissioner threatened legal action against the police to protect the public's privacy and human rights. Silkie Carlo, director of the Big Brother Watch group, described the current use of the technology as an "intrinsically Orwellian police tool."[17]

Martha Spurrier, the director of Liberty, wrote that, 'Constant surveillance leads to people self-censoring lawful behaviour. Stealthily, these measures curb our right to protest, speak freely and dissent. They shape our behaviours in ways that corrode the heart of our democratic freedoms,' adding that, 'even more perniciously, this technology is most dangerous for the people who need it the most. Technology that misidentifies women and people from ethnic minority communities disenfranchises people who already face inequality.'[18] The financial speculator and philanthropist, George Soros, known for his Open Society Foundation, described the notion of corporations like Facebook and Google colluding with the state as representing, "A web of totalitarian control… (the) likes of which not even Orwell and Huxley could imagine."[19]

Microsoft Research's Kate Crawford observed that, "Just as we are seeing a step function increase in the spread of AI, something else is happening: the rise of ultra-nationalism, right-wing authoritarianism and fascism." She added that, "This is a fascist's dream – power without accountability." AI systems often rely on data with human biases already encoded. An example highlighted by Crawford was the company Faceception which offered facial recognition technology which suggested that it could indicate character type and personality. Its material suggested that bald white men might be 'white-collar offenders', bearded middle eastern men 'terrorists', and young white women 'brand promoters.'[20]

Another disturbing problem is the potential to marry AI with pseudo-scientific ideas. For example, some in political power have been convinced by a reinvention of phrenology – the idea that a person's character and likely behaviour can be inferred from the shape of their skull which has historically been dismissed as a quack science, but also has a very bad history of abuse by totalitarian regimes. China's vice minister for science and technology has boasted that, using AI with phrenology, "we can know beforehand… who might be a terrorist, who might do something bad."[21] There are reports also of the use of the 'emotional surveillance' of workers in China, using hats and helmets fitted with sensors.[22]

Driven to distraction?

Perhaps because vehicles are so pervasive in our lives, of all applications of AI and machine learning, the subject of driverless cars has stolen headlines more than anything else. Proponents of the technology argue that it could reduce the high death rates from road crashes. But the subject is contentious because of the risk that harm from crashes may increase, and also perhaps because it is a very obvious and clear cut example of humans ceding control to 'smart' technology. Drivers are already assisted in many ways in parking, hazard awareness and even in taking their hands off the steering wheel under certain circumstances, but the fully driverless car still proves controversial.

This is partly because when things go wrong they can quickly prove to be fatal, as was the case in March 2018 when a driverless car operated by vehicle hire company Uber ran over and killed a woman who was crossing the road in Arizona.[23] Autonomous vehicles use laser sensors to build a 3-D picture of the world around them, and the one used in Uber's vehicle was reportedly the best available at the time.[24] On a wide open road it should have detected a person coming into its pathway and taken evasive action. The incident was troubling enough and shouldn't have happened, but then an investigation discovered that a key algorithm in the car's operating system had in fact identified the pedestrian, Elaine Herzberg, and actually chosen to ignore and crash into her, possibly because it did not recognise her as human. A further back-up also failed. The car had a safety driver who was meant to take control should such a system fail, but the safety driver was distracted and did not intervene. This led to the company suspending its test drives.

The fatal Uber crash might appear to be an aberration which, once corrected, will not happen again. But there are challenges and choices at the heart of the design of algorithms used in driverless vehicles which means that lethal compromise might be an inescapable design feature. There are situations in driving when a driver is forced to make decisions which can involve sacrificing the occupants of a car to protect those outside it, and vice a versa. These are moral decisions made under duress in an instant and probably on instinct. Cars conversely will act according to their coding – meaning the moral responsibility ultimately will lie outside of those in the vehicle. Who then is answerable? One study found the contradiction that while most people agreed on the principle of minimum overall harm – meaning that either those inside or outside the vehicle might be 'sacrificed' – they also said they would refuse to get into a car that was prepared to kill them.[25]

Friendly spies in the home

When a couple in Oregon, USA, discovered that a private conversation of theirs had been recorded in their home and emailed to a friend, without their knowing, by Amazon's Alexa digital assistant, they were amazed. But as strange as the incident, was the explanation given by Amazon.

It revealed a chain of causation demonstrating just how hard it is to predict the consequences of new AI-driven technology. Alexa is voice operated and, if left on, its presence easily forgotten. According to Amazon:[26]

- The Amazon Alexa Echo Dot woke up due to a word in background conversation sounding like 'Alexa'."
- Then, the subsequent conversation was heard as a 'send message' request.
- At which point, Alexa said out loud 'To whom?' At which point, the background conversation was interpreted as a name in the customer's contact list.
- Alexa then asked out loud, '[contact name], right?' Alexa then interpreted background conversation as 'right'"

Underlining the unpredictability of the technology, the Amazon spokesperson said, 'As unlikely as this string of events is... we are evaluating options to make this case even less likely.' There are numerous other accounts of products being inadvertently ordered by digital home assistants, ranging from cat food to doll's houses.[27]

Examples such as these may be disconcerting, but there are others with potentially far more disturbing implications. Toy manufacture Mattel developed and planned to introduce a smart device called Aristotle. It was designed to accompany children from their infancy to adolescence. In late 2017 Mattel cancelled the device after concerns from a range of children's interest groups, parents and lawmakers. The *New York Times*[28] reported on a petition circulated by the Campaign for a Commercial-

Free Childhood and the Story of Stuff Project, arguing that babies should not 'be encouraged to form bonds with data-collecting devices'. In addition, a joint letter from two senators complained to Mattel of the 'serious privacy concerns' created by a company building 'an in-depth profile of children and their family'. They added that, 'It appears that never before has a device had the capability to so intimately look into the life of a child.'

Other issues emerge in relation to 'friendly', connected digital devices. An investigation by consumer magazine *Which?* found that eight out of 15 domestic 'smart' gadgets – including soft toys – were vulnerable to being hacked.[29] Another study by the same journal assessed 19 smart gadgets, ranging from toothbrushes to TVs, for the information the product-makers were able to gather on consumers via them.[30] It found that 'smart, internet-connected devices you own are tracking and transmitting data on how you live your life'. The data collection being enabled amounts to a 'staggering' degree of corporate surveillance of private individuals in their homes. The devices included an upright vacuum cleaner that wanted to record audio on customers' mobile devices.

Economic impacts

In the last three years there have been several predictions of job losses due to the interaction between artificial intelligence and automation. They range from the Bank of England warning that 15 million jobs were at risk, to an Oxford University study estimating that 35 percent of current jobs in the UK would become obsolete, to an estimate by the Royal Society for the Arts (RSA) that four million private sector jobs could be lost in the next decade, with sectors most at risk including finance, accounting, transport, distribution, media, marketing and advertising.[31]

Andrew Haldane, the chief economist at the Bank of England, warned that AI threatens pervasive and prolonged unemployment.[32] He said that the upheaval in labour markets and job destruction would be unprecedented, worse than that experienced in Britain's first industrial revolution, which involved the rise of machine tools, steam power and chemical manufacturing, or the second revolution, marked by the arrival of computing.

The RSA's report, *The age of automation*, has two key conclusions:

- Achieving automation on our own terms – i.e. guaranteeing positive rather than negative outcomes – will require bold and wide-reaching policy reform.
- Who owns the machines will become an increasingly important question. This means that forms of public and workforce ownership of applications are needed to deliver broad public benefits.

Conversely, a report from the World Economic Forum, and annual gathering predominantly of the world's business elite, and based on the opinions of senior corporate executives, suggests that AI and automation will lead to a net increase of jobs. Its report, *The Future of Jobs 2018*, says the world faces the loss of 75 million jobs to the new technologies by as soon as 2022, but 'may' see 133 million new jobs similarly created in the spaces between functions carried out by AI and automation, and the real world of people, products and services. Many service economy jobs are likely to be lost in accounting, other professional services, industrial, postal and secretarial sectors. The report authors foresee that demand will increase sales, marketing and customer service sectors. According to them, however, over half of the workforce (54%) will need to be retrained for new roles.[33] The question of how the retraining of half the workforce will be financed and organised is not examined in any detail.

At certain points, economic and psychological impacts merge. As machines are upgraded with learning and immense processing abilities, will there be an inevitable human downgrade? How, for example, is the human sense of self affected if it can be beaten by machines not just in games requiring complex mathematical abilities but also in games that involve bluff? In 2017, AI beat the best human poker players.[34] Changes in the workplace in recent years have created a large category of workers termed the 'Precariat'. People with increasingly precarious livelihoods stemming from shifts that include the rise of zero-hours contracts and the conversion of what would previously been considered employees into legions of supposedly freelance suppliers (such as with delivery companies like Deliveroo and transport companies like Uber). AI and machine learning stands to intensify this dynamic in the absence of very particular interventions, and introduce new dimensions. There is an emerging debate around the legal ramifications, and the practical, moral and psychological implications of a human-AI class system. Questions hover around new notions of slavery, and at what point AI blurs distinctions with living biological consciousnesses, and then acquires rights and protections. This is neither far-fetched nor an abstract issue, given that we live in a world in which corporations – abstract structures of economic form – already possess legal personhood.

The sheer comprehensiveness of potential applications highlights how unprepared and inadequate are regulatory structures. High input industrial agriculture is often an early adopter of new technology that can have the effect of shifting the balance of power significantly between poorer, small-scale farmers and large agrochemical companies. Now with the emergence of 'digital agriculture,' concern is already high over crop ownership

and control, and liability for new applications and novel crops employing genetic manipulation. The increasing use of drones, sensors and robots in farming raises another issue related to the locking-in of unequal farming models, intensifying an often problematic business-as-usual approach. Huge volumes of data gathered through satellite, drone and other forms of monitoring can raise privacy issues and, according to the Institute for Agriculture and Trade Policy, 'be used to affect prices, insurance rates, and potentially inform land grabs in the global North and South.'[35]

As mentioned in the introduction, it is important to remember too, that in a world drowning in information, in some sectors where to function effectively means being able to make sense of big data, such as in health, AI can be applied constructively for diagnosing medical conditions. AI is already being deployed for detection, diagnosis and treatment in areas like cancer, neurology, cardiology and stroke, with many other areas of potential too.[36] And, whilst applying AI to big data could be used by judicial and security services as a tool of oppression, more informed decision-making based on a better assessment of large amounts of information is also possible.

Computer, we have a problem

"In our urge to automate, in our hurry to solve many of the world's issues, we seem to have swapped one problem for another. Everywhere you look, in the judicial system, in health care, in policing, even online shopping, there are problems with privacy, bias, error, accountability and transparency that aren't going to go away easily. Just by virtue of some algorithms existing, we face issues of fairness that cut to the core of what we are as humans, what we want our society to look like, and how far we can cope with the impending authority of dispassionate authority… Perhaps thinking of algorithms as some kind of authority is exactly where we're going wrong. Our reluctance to question the power of an algorithm has opened the door to people who wish to exploit us – a new generation of snake oil salesmen. Despite the weight of scientific evidence to the contrary, there are people selling algorithms to police forces and governments that claim to predict whether someone is a terrorist or a paedophile based on the characteristics of their face alone."[37]

Hannah Fry (2018) *Hello World: How to be Human in the Age of the Machine*

Given the very wide range of issues relating to applications of AI and machine learning, the dominant question is whether these can be effectively managed and regulated, in order that society can still derive useful benefits. Without step changes in the approach of existing regulatory mechanisms, we have to look at what governments are most likely to do in practice based on how they already operate.

The difficulty of effective regulation

A recent House of Lords report called for guiding principles for AI not to be written into law, but to be applied by existing regulatory bodies such as the Financial Conduct Authority (FCA) where finance and banking are concerned.[38] The tone of this approach follows the culture of the concerned but 'light touch regulation' which preceded the financial crisis – itself the result of complex financial products and derivatives that were only understood by very few people, even in the financial sector. It's an approach that tends to exhort and steer rather than directly regulate.

Rather than modernise a regulatory system seemingly unprepared for the implications of AI, the government's Business Secretary, Greg Clark, recently appeared to indicate a desire to weaken it further, or rather, increase its 'flexibility'. Announcing a new 'pioneer fund', Clark said it would help "unblock" 'innovation that might otherwise be hampered by red tape.'[39]

An obvious flaw in this approach is that key economic sectors such as finance and IT already exhibit the characteristics of regulatory failure and regulatory capture. Bodies charged with oversight are frequently under resourced and lack the skills and capacity to adequately handle rapid product and technological innovation. In short, at best they cannot cope, and at worst function to enable and protect business interests rather than those of the general public, whether as citizens or consumers.

The absent corporation

Various forms of automation already separate people from meaningful connection to the companies whose goods and services they use.[40] It removes direct human contact in ways which insulate corporations from feedback and direct accountability to customers. This feeds and

deepens the narrative of the 'sociopathic corporation.' [41] Resorting to the increased use of automated customer services further reduces the feedback to the owners and managers of corporations from the people and communities affected by their businesses. It also reduces the amount of real human interaction happening in a local economy.

Under what circumstances would AI benefit humanity?

The first guiding principle of the House of Lords report states that 'Artificial intelligence should be developed for the common good and benefit of humanity.' But this presupposes that all its applications must therefore be engaged in activities for the 'common good and benefit of humanity.' To deliver on that principle means that governance and ownership of those activities must explicitly be designed with public interest and benefit in mind. What would that look like in practice? It is unlikely to be either the military, or businesses in which profit maximisation under shareholder models is dominant and prioritised.

Also, in trying to decide how AI might benefit human society or not, we need thorough assessments of the potential uses and misuses. This means research on social, health, environmental and other ethical implications, which in turn means involving scientists, engineers and ethicists from a wide range of disciplines and backgrounds.

What follows in the next section is an initial contribution to such assessments from the members of Scientists for Global Responsibility – in the form of a short survey.

Survey results

The membership of Scientists for Global Responsibility is drawn from diverse fields of science and social science, engineering and technology. Around half are drawn from the natural sciences, such as physics, chemistry and biology, with the next largest group being in engineering and information technology, and 10% in the social sciences. The membership ranges from students to the retired, and includes researchers and lecturers, those working in academia, the public sector, business and other non-governmental organisations.[42]

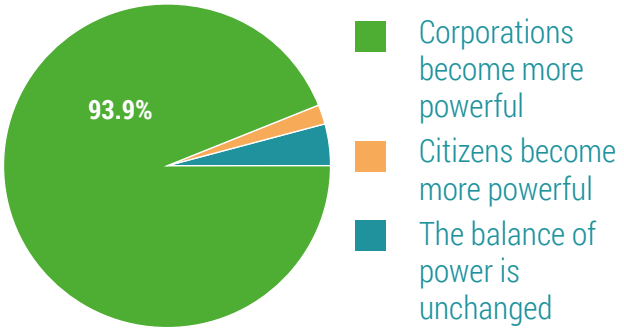
As professionals for whom the emergence of AI, automation and machine learning will have direct implications, we conducted a poll amongst them and our wider supporter base to gather an expert sounding on the issues.[43]

The results represent something of a wake-up call over the level of concern among this group about how the development and introduction of the technology and applications is being handled. We asked 8 specific questions and one open question in which respondents were invited to say what, for them, represented the 'biggest issue' with regard to AI. The survey was introduced in this way:

> *Artificial intelligence – autonomous learning and/or decision making by machines - is increasingly being planned for and deployed - on the road with driverless cars, on the battlefield with autonomous weapons, and in the home and workplace with digital assistants. How do you think this is likely to change the world and how we live in it?*

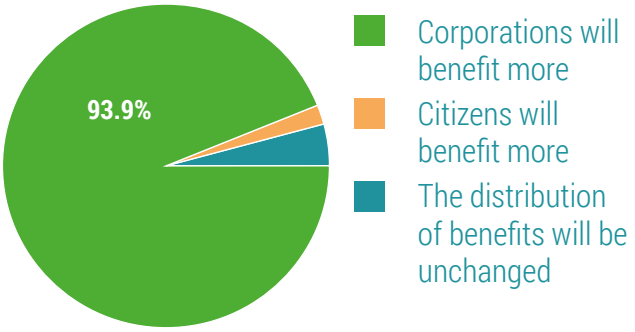
Question 1 - Artificial intelligence is increasingly being used by providers of goods and services. In the balance of power between corporations and citizens what do you think will be the most likely outcome?

Respondents were given three options: corporations become more powerful, citizens become more powerful, and that the balance of power between citizens and corporations would be unchanged. A huge majority, 94%, believed that AI would give corporations more power over citizens. Just one percent thought the reverse.



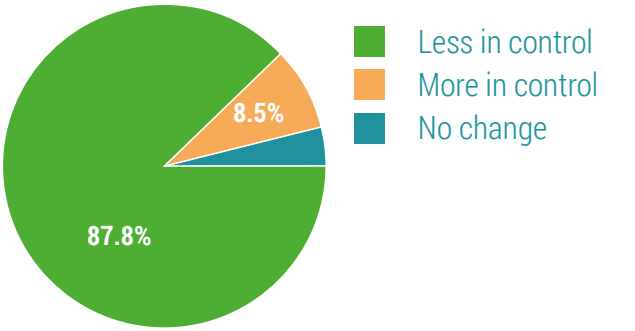
Question 2 - As a consequence of the above, how will this most likely effect the distribution of economic benefits?

The second question was designed to find out what people thought would be the real world consequences of their answers to the first. Again, people were given three options: corporations will benefit more, citizens will benefit more, and the distribution of benefits will be unchanged. The answers to this question mirrored the first, equally strongly. Again 94% thought that corporations were likely to benefit more from the implementation than citizens, with only one percent thinking the reverse.



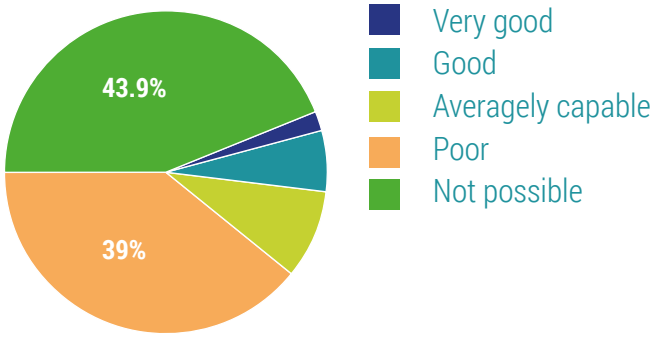
Question 3 - Does the prospect of greater use of artificial intelligence make you feel: more in control of your life, less in control of your life or, no change?

In our third question we asked people how the prospect of greater deployment of AI made them feel in terms of the crucial issue about the level of agency we believe we have. Or, in other words, how much control we feel we have over our own lives. Not only has this question profoundly shaped politics in recent years, but it is also known to be an important dimension of our well-being. Once again, the results revealed a very strong outcome. Even among a group containing a very probably disproportionate level of technical expertise due to the nature of the membership, 88% responded that the prospect made them feel less in control of their lives, while just 4% thought the opposite, with 8% citing no change.



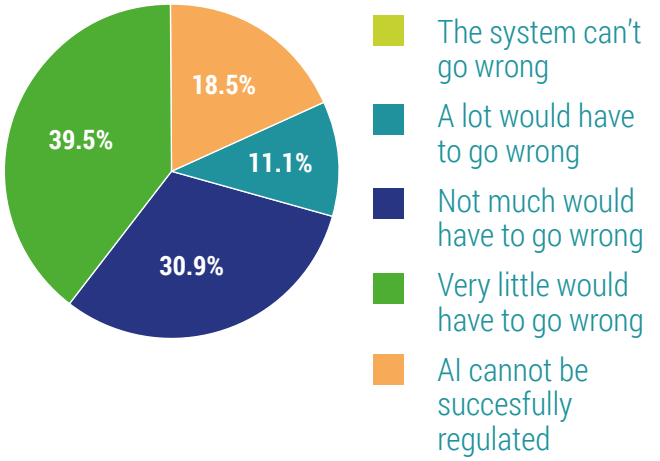
Question 4 - How good do you think those designing and introducing artificial intelligence are at predicting the fullest likely range of its consequences?

In question 4 we moved on to ask people how confident they were about the ways in which AI was being developed, and how confident they were that developers were effectively in the driving seat of applications that were emerging. This time more options were given in order for people to be able to rank their relative levels of confidence. Answers indicated most strongly that current abilities to predict consequences were either low or absent. Combined, 83% said the ability to predict consequences was either 'poor' or 'not possible'. Only 8% said that AI developers were either 'good' or 'very good' at predicting consequences.



Question 5 - Given current regulations and regulators how much do you think would have to go wrong for artificial intelligence to cause significant harm?

In question 5 was asked people how prepared was the current regulatory framework to manage the multiple issues to do with AI. Stories of AI-assisted hacking of social media, and of news distribution targeted at influencing elections, has put the issue high on the political agenda, suggesting action will be a priority for regulators. Few, however, seemed convinced that current regulatory frameworks and responses were adequate or fit for the task in hand. 71% thought that 'not much' or 'very little' would have to go wrong for AI to cause significant harm. An additional 19% said that it was not possible for AI to be successfully regulated.



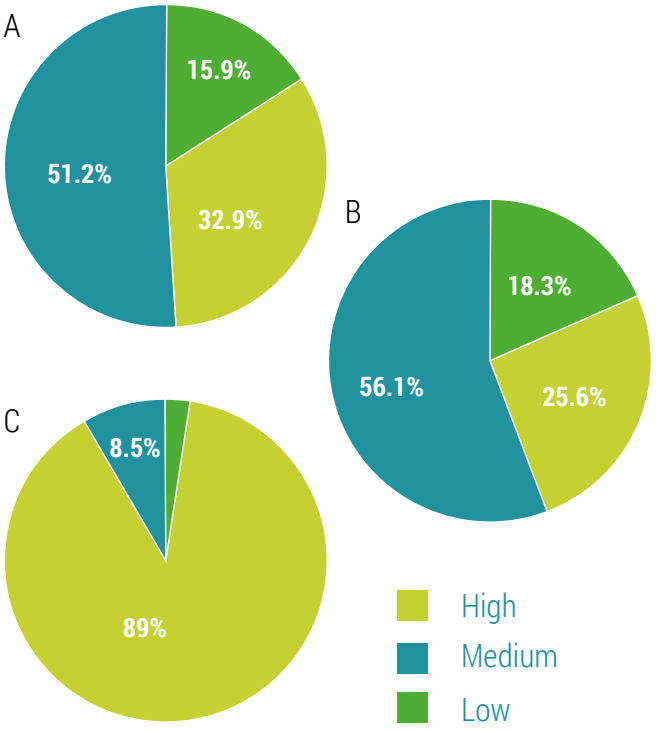
Question 6

A. How would you rate the chances of things going badly wrong with artificial intelligence in driverless cars?

B. How would you rate the chances of things going badly wrong with artificial intelligence in digital assistants?

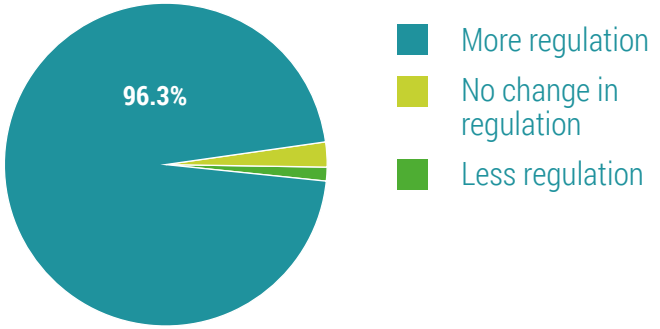
C. How would you rate the chances of things going badly wrong with artificial intelligence in autonomous weapons?

Our next question we broke down into three parts to ascertain people's relative confidence concerning the application of AI in three areas which have proven differently controversial. We asked how respondents would rate the chances of things going badly wrong in relation to driverless cars, digital assistants and autonomous weapons. Responses tended to reflect the broader perception of threat from the type of technology being considered. With driverless cars, 84% thought there was a medium or high chance of things going badly wrong. For digital assistants the comparable response was 82%. Respondents were most concerned that there was a chance of things going badly wrong where autonomous weapons were concerned. Here 97% thought that to be the case.



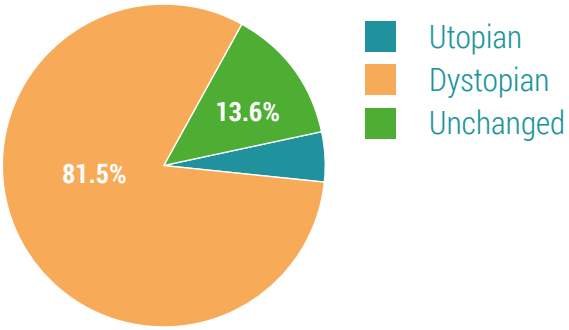
Question 7 - Do you think artificial intelligence needs: more regulation, no change in regulation or, less regulation?

Question 7 was straightforward and designed to determine what people thought should be the future trajectory of regulation of AI. We wanted to know whether people thought there should more checks and balances, if the balance was about right, or whether regulation was holding back an important new development. An unambiguous 96% thought that AI needed more regulation.



Question 8 - Do you think artificial intelligence will create a future which tends more towards being: utopian, dystopian, or unchanged?

We wanted to know what kind of future people thought we were heading towards under current circumstances with the assistance of AI. The idea that there is a technological fix for many things is often attractive. If an intelligent machine or application can solve our problems, it bears the promise that other things won't have to change too much. Technological optimism and the techno-fix as an approach is common and attractive. And people for whom science and technology is their livelihood might see advantage in it. For that reason we asked people whether they thought that AI was likely to set us on a positive course of improvement or more likely to head the other way. Among respondents, 14% thought the future would be unchanged, and 5% were persuaded that a utopian future awaited us. But 82% thought that AI was most likely to create a dystopian future.



Question 9 - What do you think is the biggest issue to do with artificial intelligence, and why?

Our final substantive question was an open one in which we asked people about what they thought was the most important issues to do with AI, and to explain why they thought so. Respondents could raise and say whatever they wanted without any particular guidance. Below is a selection of answers from respondents, which are indicative of the range of replies we received:

- *Loss of control, because we are poor at imagining emergent consequences. Also, because the more successful is the creation of artificial intelligence, the greater is the likelihood that things will happen unpredictably and beyond our control*
- Autonomous military systems, including weapons. Such technologies create and/or increase the potential for war by accident - including nuclear war.
- *The potential to increase inequality - a terrible problem made worse*
- Use of AI for life-changing decisions such as offering mortgages, medical decisions, granting parole, "pre-crime" - potentially arresting people "preventatively"
- *The control & design of technology being chiefly funded and thus controlled by large unaccountable organisations such as Google, Alphabet, Microsoft, CA etc*
- Its design criteria prioritising profit over social benefits
- *Independent intelligence will develop independent goals, without any relation to human values, and may well be able to carry them out.*
- People will not have the capacity/engagement to understand the implications of their interactions with AI, and so will not adequately consider issues such as consent, privacy. Corporations will not, either through ineptitude or lack of caring and regulation, give adequate protection to the vulnerable in planning their use of AI. Thus the big issue which will emerge is the unintended consequence of sleepwalking into a situation where the masses are 'catered to' (albeit imperfectly) and the vulnerable are severely compromised.
- *Introduced too quickly without thought for the consequences, then taking too long to regulate.*
- Intellectual: The undermining of creative thought processes and critical analysis, and the abdication of collective responsibility through the adoption of a quick-fix, tick-box mentality.
- *There are potentially many benefits if the chain of accountability and distributed ethics are adequately tackled. However, I am a member of the International Committee for Robot Arms Control which is devoted to achieving a ban on all autonomous weapons where there is no meaningful human control.*
- Societal collapse due to destruction of employment opportunities for all but a privileged few
- *Paradoxically, it will not match up to its claims but it will never-the-less significantly change our culture. People will believe the claims to be 100% valid/feel there is little that can be done and give away too much power and responsibility to A.I.*

Corrupted code

Conclusion and questions

In our small survey, the expert membership of Scientists for Global Responsibility, a mixture of engineers, technologists, researchers and scientists, raise numerous deep concerns about the prospects of AI and machine learning, and issue a call for far greater oversight and assessment.

The technologies and applications are entering a human society which is consuming natural resources at unsustainable rates and, in cases such as climate change, already passing critical planetary ecological boundaries. It is a society that, in many places, is over-weaponised and characterised by corrosive inequality. Without radical change, new forms of automation and artificial intelligence will be owned by and serve the agenda of the people, corporations and interests who preside over the world as it is, greatly increasing their already significant power. Far greater asymmetry, imbalance and destructive potential then becomes not just possible, but probable.

These issues speak to our core concerns about 'responsibility' in science. As a very notion, responsibility speaks to the holding of human consciousness, identity and moral agency. At a fundamental level, the introduction of AI seeks to remove the immediate need for humans to exercise direct responsibility – outsourcing it via AI. There are already multiple instances in which other forms of conventional economic outsourcing has failed, in contexts ranging from transport and care homes to security, and a wide range of sectors. This bodes badly for the more philosophically challenging outsourcing of the responsibilities involved in making choices that have material impacts on the lives of people. It runs the risk of embedding and institutionalising a form of sociopathy. One of the most emotive dimensions of the recent Brexit debate has been the notion of 'taking back control.' It is ironic therefore that through the rapid development and deployment of AI and machine learning, we appear to be handing over control of fundamental aspects of our lives, politics and economies with poor understanding of the full implications and consequences of doing so.

It is darkly ironic that the House of Commons select committee on Digital, Culture, Media and Sport (DCMS) saw, in their investigation into the Facebook and Cambridge Analytica, evidence of:

> *"'fake news', created for profit or other gain, disseminated through state-sponsored programmes, or spread through the deliberate distortion of facts, by groups with a particular agenda, including the desire to affect political elections [and] …evidence of Russian state-sponsored attempts to influence elections in the US and the UK through social media, of the efforts of private companies to do the same, and of law-breaking by certain Leave campaign groups in the UK's EU Referendum in their use of social media"*[44]

One consequence that is easier to foresee is that, without specific interventions designed to produce different outcomes, the use of AI, further automation and machine learning by already powerful economic and political actors, will further entrench and strengthen their positions. It seems reasonable to conclude that this will lead to more 'business as usual', locking in and intensifying existing patterns of consumption, commodification, inequality, and the concentration of wealth and power.

At the same time, AI could allow the powerful to distance themselves further from taking responsibility for their actions, and make accountability more difficult. It also threatens an assault on human relationships based on the exercise of trust, responsibility, accountability and moral judgement.

Rather than proposing actual legislation, the House of Lords set out guiding principles for the regulation of AI. The implication, however, is that if the conditions they outline as necessary do not apply, then AI applications should not be permitted. Their guidelines are that:[45]

- Artificial intelligence should be developed for the common good and benefit of humanity.
- Artificial intelligence should operate on principles of intelligibility and fairness.
- Artificial intelligence should not be used to diminish the data rights or privacy of individuals, families or communities.
- All citizens should have the right to be educated to enable them to flourish mentally, emotionally and economically alongside artificial intelligence.
- The autonomous power to hurt, destroy or deceive human beings should never be vested in artificial intelligence.

Were these guidelines to be followed effectively they would seem, at the very least, to rule out the deployment of autonomous weapons or the use of AI by security or

police forces as long as miscarriages of justice resulting from its use were likely.

But, regardless, it would be very hard to argue that any of these conditions could currently be guaranteed, which would in turn argue for a pause to allow regulators to catch-up before further dissemination of AI is permitted, especially in the area of sensitive applications, such as state security and defence.

The purpose of this briefing is to survey and summarise a number of critical issues relating to the development and deployment of AI, automation and machine learning that need further scrutiny. It is not its purpose, and neither are we in a position, to make detailed, comprehensive recommendations for regulation. However, enough information is already in the public domain for us to make two urgent calls.

First, with already the backing of large numbers in the AI and related fields, there should be a ban on the development and deployment of autonomous weapons, and the UK government should support this through the UN Convention on Certain Conventional Weapons.

Second, in order to create the conditions for an effective regulatory framework, which currently does not exist, 20% of AI research and development budgets should be spent on assessments of the potential benefits and harms of innovations and applications in the field, including suggesting necessary measures to prevent harm.

We began with the question 'AI – how little has to go wrong?' The answer seems to be not much before there are serious consequences. Lethal mistakes have already been made. The challenge now is to create the conditions in which things are most likely to go right. That is a task for responsible science and society.

Endnotes

1 Harvard Business Review (10 May 2018) Promising AI Applications in Health Care, https://hbr.org/2018/05/10-promising-ai-applications-in-health-care; Jiang F, Jiang Y & Zhi H *et al* (2017) Artificial intelligence in healthcare: past, present and future, *Stroke and Vascular Neurology*, 2:doi:10.1136/svn-2017-000101.

2 Mustafa Suleyman, Co-founder and Head of Applied AI at DeepMind, speaking on BBC Radio 4 (6 September, 2018).

3 *The Future of Life Institute*, Autonomous weapons: an open letter from AI & robotics researchers. https://futureoflife.org/open-letter-autonomous-weapons/

4 See for example, Lindqvist, S (2001) *A History of Bombing*, Granta.

5 *The Future of Life Institute* (4 April 2018) AI and Robotics Researchers Boycott South Korea Tech Institute Over Development of AI Weapons Technology. https://futureoflife.org/2018/04/04/ai-and-robotics-researchers-boycott-kaist/

6 *The Economist* (25 April 2018) The Future of War.

7 Simms, J (2017) *Over by Christmas?* http://www.historynet.com/over-by-christmas.htm

8 Tegmark, M (2017) *Life 3.0*. Allen Lane.

9 Futter, A (2018) *Cyber Threats and Nuclear Weapons*, Oxford Research Group. https://www.oxfordresearchgroup.org.uk/blog/cyber-threats-and-nuclear-weapons

10 *Washington Post* (2 June 2012) Stuxnet was work of U.S. and Israeli experts, officials say.https://www.washingtonpost.com/world/national-security/stuxnet-was-work-of-us-and-israeli-experts-officials-say/2012/06/01/gJQAlnEy6U_story.html?noredirect=on&utm_term=.dcdf60fe9456

11 CBC/Thomson Reuters (16 February 2015) NSA hid spying software in hard drive firmware, report says; Bloomberg (04/10/2018) The Big Hack – how China used a tiny chip to infiltrate America's top companies. https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies

12 Simms, A (2007) *Tescopoly: how one shop came out on top and why it matters*, Constable

13 *The Guardian* (16 April 2018) Cambridge Analytica scandal 'highlights need for AI regulation'.

14 *The Observer* (28 July 2018). https://www.theguardian.com/technology/2018/jul/28/dcms-report-fake-news-disinformation-brexit-facebook-russia

15 House of Lords (16 April 2018) *Artificial Intelligence Committee, AI in the UK: ready, willing and able?* Report of Session 2017–19.

16 *The Independent* (15 May 2018) Information commissioner threatens legal action against police using 'dangerous and inaccurate' facial recognition technology.

17 *Ibid*.

18 Martha Spurrier (16 May 2018) Facial recognition is not just useless. In police hands, it is dangerous.

19 *AFP* (26 January 2018) Soros to Google and Facebook: 'Your days are numbered'. https://phys.org/news/2018-01-soros-google-facebook-days.html

20 Kate Crawford (12 March 2017) *DARK DAYS: AI and the Rise of Fascism*. https://schedule.sxsw.com/2017/events/PP93821

21 *Financial Times* (23 July 2017) China seeks glimpse of citizens' future with crime-predicting AI … https://www.ft.com/content/5ec7093c-6e06-11e7-b9c7-15af748b60d0

22 Business Insider (1 May 2018) China is monitoring employees' brain waves and emotions. https://www.businessinsider.com/china-emotional-surveillance-technology-2018-4

23 *The Times* (14 May 2018) Uber's Driverless Car 'Chose To Hit' Woman Who Died. https://www.driving.co.uk/news/ubers-driverless-car-chose-hit-woman-died/

24 Aarian Marshall (31 March 2018) The Uber crash won't be the last shocking self-driving death. *Wired.* https://www.wired.com/story/uber-self-driving-crash-explanation-lidar-sensors/

25 Bonnefron *et al* (2016) *The social dilemma of autonomous vehicles, Computers and Society*. https://arxiv.org/abs/1510.03346

26 BBC News (24 May 2018) *Amazon Alexa heard and sent private chat.* https://www.bbc.co.uk/news/technology-44248122

27 GeekWire (2017) https://www.geekwire.com/2017/girl-6-accidentally-orders-170-dollhouse-4-pounds-cookies-using-amazons-alexa/

28 *New York Times* (2017) Mattel Pulls Aristotle Children's Device After Privacy Concerns. https://www.nytimes.com/2017/10/05/well/family/mattel-aristotle-privacy.html

29 *Which?* (23 June 2017) Could your smart home be hacked? https://www.which.co.uk/news/2017/06/could-your-smart-home-be-hacked/

30 *Which?* (1 June 2018) Investigation reveals 'staggering' level of smart home surveillance. https://www.which.co.uk/news/2018/06/which-investigation-reveals-staggering-level-of-smart-home-surveillance/

31 RSA (2017) *The Age Of Automation: Artificial Intelligence, Robotics And The Future Of Low-Skilled Work*. https://www.thersa.org/discover/publications-and-articles/reports/the-age-of-automation

32 *Financial Times* (20 August 2018) Bank of England economist says disruption will be more intense than industrial revolution

33 WEF (2018) *The Future of Jobs Report 2018*. http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf

34 *Wired* (1 January 2017) Inside Libratus, The Poker Ai That Out-Bluffed The Best Humans. https://www.wired.com/2017/02/libratus/

35 Karen Hansen-Kuhn (2018) Ag Tech Takeback – Debate by ARC2020 and IATP. http://www.arc2020.eu/agtechtakeback-debate/

36 Jiang F, Jiang Y & Zhi H *et al* (2017) Artificial intelligence in healthcare: past, present and future, *Stroke and Vascular Neurology*, 2:doi:10.1136/svn-2017-000101.

37 Hannah Fry author of *Hello World: How to be Human in the Age of the Machine*, speaking on BBC Radio 4 (15 September 2018).

38 House of Lords (16 April 2018) *Artificial Intelligence Committee, AI in the UK: ready, willing and able?* Report of Session 2017–19.

39 *The Guardian* (5 October 2018) 'Robolawyers' among futuristic ideas backed by £10m innovation fund. https://www.theguardian.com/business/2018/oct/05/business-secretary-unveils-10m-fund-for-innovation?CMP=Share_iOSApp_Other

40 New Weather Institute (2017) *The Absent Corporation.* http://www.newweather.org/2017/02/28/how-companies-are-hiding-from-customers/

41 Joel Bakan (2005) *The Corporation: the pathological pursuit of profit and power.*

42 At the time of writing, the organisation had 750 members, of whom 16% were 'associates' who have concerns about ethical issues in science, design and technology, but not necessarily a related professional background.

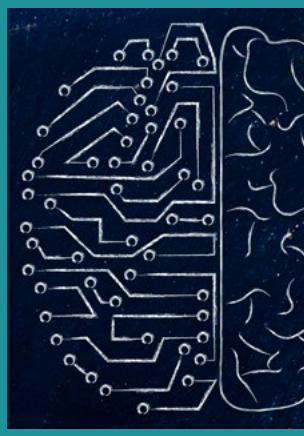
43 There were 82 respondents to the poll, 85% of which were actual members, representing just under one in ten of the membership, and 15% broader supporters.

44 House of Commons (29 July 2018) *Digital, Culture, Media and Sport Committee Disinformation and 'fake news': Interim Report* Fifth Report of Session 2017–19.

45 House of Lords (16 April 2018) *Artificial Intelligence Committee, AI in the UK: ready, willing and able?* Report of Session 2017–19.

Scientists for Global Responsibility

Scientists · Architects · Engineers · Technologists



Published by Scientists for Global Responsibility, November 2018.

Scientists for Global Responsibility promotes responsible science, design and technology, and is part of the campaign awarded the Nobel Peace Prize in 2017.

SGR is an independent UK-based membership organisation of hundreds of natural scientists, social scientists, engineers, IT professionals and architects. We promote science, design and technology that contribute to peace, social justice, and environmental sustainability.

SGR's work is focused on four main issues: security and disarmament; climate change and energy, including nuclear power; who controls science and technology? and emerging technologies.

SGR is affiliated to the International Network of Engineers and Scientists for Global Responsibility (INES).

Please support SGR's work by becoming a member - for details, see: www.sgr.org.uk

FSC Printed on Evolution Indigo FSC certified, recycled paper by Precision Printing.