Chatham House Report

M. L. Cummings, Heather M. Roff, Kenneth Cukier, Jacob Parakilas
and Hannah Bryce

Artificial Intelligence and International Affairs

Disruption Anticipated

#CHAI

CHATHAM
HOUSE
The Royal Institute of
International Affairs

Chatham House Report

M. L. Cummings, Heather M. Roff, Kenneth Cukier, Jacob Parakilas and Hannah Bryce
International Security Department & US and the Americas Programme | June 2018

Artificial Intelligence and International Affairs

Disruption Anticipated

CHATHAM HOUSE
The Royal Institute of International Affairs

Chatham House, the Royal Institute of International Affairs, is an independent policy institute based in London. Our mission is to help build a sustainably secure, prosperous and just world.

The Royal Institute of International Affairs

Chatham House
10 St James's Square
London SW1Y 4LE
T: +44 (0) 20 7957 5700
F: + 44 (0) 20 7957 5710
www.chathamhouse.org

Charity Registration No. 208223

ISBN 978 1 78413 212 5

A catalogue record for this title is available from the British Library.

Printed and bound in Great Britain.

The material selected for the printing of this report is manufactured from 100% genuine de-inked post-consumer waste by an ISO 14001 certified mill and is Processed Chlorine Free.

FSC

Typeset by Soapbox, www.soapbox.co.uk

Cover image: Self-driving cars during a road test on 22 March 2018 in Beijing, China. Beijing's traffic authority has issued temporary number plates to self-driving cars developed by the Chinese search engine Baidu.
Copyright © Anadolu Agency/Contributor/Getty Images

Contents

Executive Summary

For all of human history, politics has been fundamentally driven by conscious human action and the collective actions and interactions of humans within networks and organizations. Now, advances in artificial intelligence (AI) hold out the prospect of a fundamental change in this arrangement: the *idea* of a non-human entity having specific agency could create radical change in our understanding of politics at the widest levels.

Not least because of the influence of literature, cinema and television, popular thinking about AI can tend towards the fanciful. Fictional, apocalyptic depictions of war between humans and robots have influenced breathless coverage of sometimes relatively minor AI developments. Periodically, too, leading figures in the fields of science and technology have issued stark warnings that AI may pose an existential threat to human life. Together, these have given rise to a perception among the general public that a new form of intelligence that exceeds human intelligence is just around the corner – or even with us already.

Humans and limited forms of AI already coexist: AI technology helps us to navigate, to translate text and to find cheap flights, to give just a few examples; and – notwithstanding its known flaws and limitations – it looks set to be emblematic of a radically transformed future.

But the more extreme ideas of what advances in AI may mean for how humans live, work and engage is far distant from the current reality. The nature of AI in 2018 – and very likely for the foreseeable future – is somewhat mundane. Indeed, the field is seeing relatively minor advancements that bring specific practical benefits in identified areas, rather than AI with general application. Services such as Google Translate, for instance, are undoubtedly useful, but the increased efficiencies created by such services do not yet hold out the prospect of noticeably changing the power balance at the international level. A truly non-human intelligence would likely do so, but a constructed system capable of operating on a comparable level to a human brain – an artificial general intelligence, or AGI – would require a broad-based advancement in every aspect of the field: hardware, software, and even our understanding of what cognition actually is.

Technological change does not have to be dramatic or sudden to create meaningful shifts in power balances or social structures

The more prosaic advancements are not insignificant, however. Technological change does not have to be dramatic or sudden to create meaningful shifts in power balances or social structures. Indeed, focusing on the distant prospect of dramatic change may well distract from developing a more nuanced understanding of slower and subtler, but equally significant, changes.

This Chatham House report examines some of the challenges for policymakers, in the short to medium term, that may arise from the advancement and increasing application of AI. It is beyond the scope of the report to offer a fully comprehensive set of predictions for every possible ramification of AI for the world. Significant areas not addressed here – including medicine, public health and law – might be fundamentally transformed in the next decades by AI, with considerable impacts on the processes of the international system. Furthermore, towards the end of the process of compiling the report, public attention has increasingly turned to the possibility of AI being used to support disinformation campaigns or interfere in democratic processes. We intend to focus on this area in follow-up work.

This report does not attempt to offer specific predictions for the progress of discrete technological avenues, or proposals for specific avenues of technological development. Rather, it draws together strands of thinking about the impact that AI may have on selected areas of international affairs – from military, human security and economic perspectives – over the next 10 to 15 years.

The report sets out, first, a broad framework to define and distinguish between the types of roles that artificial intelligence might play in policymaking and international affairs: these roles are identified as analytical, predictive and operational.

In **analytical** roles, AI systems might allow fewer humans to make higher-level decisions, or to automate repetitive tasks such as monitoring sensors set up to ensure treaty compliance. In these roles, AI may well change – and in some ways it has already changed – the structures through which human decision-makers understand the world. But the ultimate impact of those changes is likely to be attenuated rather than transformative.

Predictive uses of AI could have more acute impacts, though likely on a longer timeframe. Such employments may change how policymakers and states understand the potential outcomes of specific courses of action. This could, if such systems become sufficiently accurate and trusted, create a power gap between those actors equipped with such systems and those without – with notably unpredictable results.

Operational uses of AI are unlikely to fully materialize in the near term. The regulatory, ethical and technological hurdles to fully autonomous vehicles, weapons and other physical-world systems such as robotic personal assistants are very high – although rapid progress towards overcoming these barriers is being made. In the longer term, however, such systems could radically transform not only the way decisions are made but the manner in which they are carried out.

The report then turns to examine the near-term implications of AI applications in the military, human security and economic fields. Missy Cummings, looking at the military sector, concludes that truly autonomous weapons systems are still some distance away: a combination of operational and doctrinal issues have largely prevented their adoption hitherto, although remotely operated vehicles are increasingly prevalent for some applications such as aerial and undersea reconnaissance. She argues that a significant shift in gravity is under way between traditional defence industries and the non-defence technology industry, with implications for how military systems are designed and acquired.

Heather Roff argues that AI does have positive implications for human security, but that unlocking progress means first understanding the roles in which AI can be put to positive use – and, critically, understanding the difference between using data (which machines can sort effectively) and knowledge (which humans remain far better at). She concludes, furthermore, that in order to fully reap the potential benefits of AI in the realm of human security, proactive steps need to be taken to ensure equality of access to technology.

Kenn Cukier makes the case that AI is likely to reshape what work looks like, but that it is unlikely to fundamentally change underlying economic power structures. Artificially intelligent systems – both those employed in operational roles (like autonomous vehicles) and those in analytical and predictive roles – are likely, in his view, to create significant wealth, but the distribution of that wealth will not inherently become more equal for humans.

Building a framework for better managing the rise of artificially intelligent systems in the near term might also reinforce the process of mitigating longer-term risks

AI technology may have profound impacts on economic and geopolitical power balances, but it will require clarity of purpose to ensure that it does not simply serve to reinforce existing inequities. Building a framework for better managing the rise of artificially intelligent systems in the near term might also reinforce the process of mitigating longer-term risks. To this end, the report makes the following recommendations for governments and international non-governmental organizations, which will have a particularly important role in developing and advocating for new ethical norms:

- **In the medium to long term, AI expertise must not reside in only a small number of countries – or solely within narrow segments of the population.** Governments worldwide must invest in developing and retaining home-grown talent and expertise in AI if their countries are to be independent of the dominant AI expertise that is now typically concentrated in the US and China. And they should work to ensure that engineering talent is nurtured across a broad base in order to mitigate inherent bias issues.
- **Corporations, foundations and governments should allocate funding to develop and deploy AI systems with humanitarian goals.** The humanitarian sector could derive significant benefit from such systems, which might for example decrease response times in emergencies. Since AI for humanitarian purposes is unlikely to be immediately profitable for the private sector, however, a concerted effort needs to be made to develop them on a not-for-profit basis.
- **Understanding of the capacities and limitations of artificially intelligent systems must not be the exclusive preserve of technical experts.** Better education and training on what AI is – and, critically, what it is not – should be made as broadly available as possible, while understanding of underlying ethical and policy goals should be a much higher priority to those developing the technologies.
- **Developing strong working relationships, particularly in the defence sector, between public and private AI developers is critical, as much of the innovation is taking place in the commercial sector.** Ensuring that intelligent systems charged with critical tasks can carry them out safely and ethically will require openness between different types of institutions.
- **Clear codes of practice are necessary to ensure that the benefits of AI can be shared widely while its concurrent risks are well managed.** In developing these codes of practice, policymakers and technologists should understand the ways in which regulating artificially intelligent systems may be fundamentally different from regulating arms or trade flows, while also drawing relevant lessons from those models.
- **Particular attention must be paid by developers and regulators to the question of human–machine interfaces.** Artificial and human intelligence are fundamentally different, and interfaces between the two must be designed carefully, and reviewed constantly, in order to avoid misunderstandings that in many applications could have serious consequences.

1. Introduction: Artificial Intelligence and International Politics

Jacob Parakilas and Hannah Bryce

Across the entire spectrum of human behaviours, politics may be one of the most difficult activities to automate. Politics, as it is commonly understood (as the mechanism by which competing objectives are weighed against each other[1]), is an inherently complex task that reflects the complexity of human behaviour on both an individual and a mass scale. This becomes all the more true at the level of international relations.

It seems unlikely that human-level artificial intelligence (AI) – so-called artificial general intelligence (AGI) – will emerge in the near future. Even if the state of the art advances much more quickly than anticipated, there is considerable resistance to the idea of turning responsibility over to machines. This can be seen playing out notably in the present-day debates over autonomous cars and robotic weapons systems. In this context, therefore, it is difficult to envisage a world in which the decision-making elements of politics are outsourced to machines in their entirety without also imagining a completely different, speculative world of the future.

While the probability of a robotic (in the literal sense) president or foreign minister seems very distant, this does not mean that AI will not affect international politics in significant ways. Its impacts are likely to be more diffuse, and subtle. They are likely to be manifested through changes in the ways in which human decision-makers are informed, while not extending so far as to allow AI to take decisions. Consideration of the application of AI in international affairs appropriately includes what the structures that support decision-makers look like, and the speed with which critically significant decisions are made.

It seems safe to assume that artificially intelligent systems will not replace humans at the top level of decision-making. But they will be an increasingly significant part of the context in which human decision-makers operate

In short, it seems safe to assume that artificially intelligent systems will not replace humans at the top level of decision-making. But they will be an increasingly significant part of the context in which human decision-makers operate. This evolution presents both huge opportunities and substantial risks, so considering the potential impacts at an early stage is critical.

How might AI fit into international relations?

Ultimately, AGI might be capable of executing any cognitive or operational task for which human intelligence is currently necessary. But given the likelihood that such AGI will take decades, or perhaps centuries, to develop, current analysts and policymakers might reasonably focus chiefly on consideration of tasks that may be assigned to AI in the relatively near term.

Such tasks will depend heavily on the capabilities of AI. Machines are of course capable of processing enormous amounts of data exceptionally quickly. They can also store and access far greater amounts of data than can a human mind. With the correct software, a machine can recognize patterns in data much more quickly and accurately than any human can. But machines also operate along a very limiting set of parameters – where a human child can instinctively recognize a cat from any angle, a computer (even after sorting thousands of cat images) can be flummoxed by seeing a cat whose face is temporarily hidden from view.[2]

[1] Weber, M. (1958), 'Politics as a Vocation' in Gerth, H. H. and Wright Mills, C. (eds) (trans) (1958), *From Max Weber: Essays in Sociology*, New York: Oxford University Press.
[2] Le, Q. V., Ranzato, M.A., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J. and Ng, A. Y. (2012) 'Building High-level Features Using Large Scale Unsupervised Learning' in *Proceedings of the 29th International Conference on Machine Learning*, https://arxiv.org/abs/1112.6209.

Consider, for example, the case of the first fatal crash of a vehicle being operated in ‘self-drive’ mode: a Tesla Model S, which in May 2016 drove at full-speed into the side of a truck; its human driver was killed in the collision. According to investigators, the car’s sensors were confused by sunlight reflecting off the white paint of the truck’s trailer, which it was unable to distinguish from the sky. The system neither braked nor warned the human driver of the impending collision. Investigators concluded that the ultimate responsibility lay with the human driver, whose failure to properly oversee the operation of the vehicle led to the accident.[3]

Machines and humans have different capabilities, and, equally importantly, make different mistakes based on fundamentally divergent decision-making architectures

Regardless of the legal responsibility, it is hard to imagine a human driver making that particular mistake. By the same token, it is impossible to imagine an AI system making the kinds of error that human drivers do frequently and with often horrific consequences – such as driving while tired, distracted or drunk. Machines and humans have different capabilities, and, equally importantly, make different mistakes based on fundamentally divergent decision-making architectures.[4]

One frequently mooted solution is to combine humans and machines in teams that allow them to operate in complementary fashion. Provided a suitably useful interface can be set up to mediate between the two – in itself a not insignificant obstacle – humans working with machines might combine the strengths of both while avoiding the pitfalls associated with either. Teams following this arrangement, dubbed ‘centaurs’ after the half-man, half-horse of Greek mythology, have already been trialled for military use by the US, and seem likely to be the goal for the near future at least.

The key with these blended approaches is a clear delineation of responsibility. Unless the human–machine interfaces – both at the individual operator level and higher up the chain – are designed and tested with extreme care, the teaming brings with it completely new terrain in which mistakes can be made and where responsibility can slip through the cracks. And as the scope of responsibility increases, so too do the consequences.

Explored next are the three categories where AI is likely to be used in a particularly instrumental way in international politics and policymaking.

Analytical roles

Artificially intelligent systems are already found in analytical roles, combing through large datasets and deriving conclusions based on pattern recognition. These are precisely the ‘dull’ tasks (of the ‘dull, dirty and dangerous’ formulation[5]) that are generally regarded as the highest priority for automation.

[3] Boudette, N. E. (2017), ‘Tesla’s Self-Driving System Cleared in Deadly Crash’, *New York Times*, 19 January 2017, https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html.
[4] This incident was followed, in March 2018, by the first case of an autonomously operated vehicle killing a pedestrian, when a Volvo XC90 operated as part of a development programme by ride-hailing company Uber struck and killed a woman in Tempe, Arizona. See Levin, S. and Wong, J. C. (2018), ‘Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian’, *The Guardian*, 10 March 2018, https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe.
[5] *The Economist* (2011), ‘Flight of the drones: Why the future of air power belongs to unmanned systems’, 8 October 2011, http://www.economist.com/node/21531433.

The possibilities of such intelligences are iterative: they will increase with the spread and growth in capability of artificially intelligent software. In Chapter 2 of this report, Missy Cummings explores in fuller detail the extent to which AI is being used within weapons systems, and the issues and challenges that this presents. But a few possible missions or roles can already be imagined without too much difficulty.

Monitoring the outputs of sensors set up to verify compliance with, for instance, a nuclear, chemical or biological arms control treaty might well be a deadening job for human analysts – albeit one that would require significant specialist training and experience. By contrast, a machine learning system[6] set up to do the same would never tire or grow bored of its task. And while it might (especially in the process of learning) flag a large number of 'false positives', by assigning human specialists to oversee and correct it the system's accuracy would quickly increase.

In a similar, albeit less dramatic fashion, artificially intelligent processes might be very useful for optimizing the procedures of the more mundane aspects of political exchange. Given the increasing amount of real-time data generated by industrial and commercial operations (through what is often referred to as the 'internet of things', or IoT), it is not difficult to imagine artificially intelligent systems monitoring trade and feeding into decision-making processes around economic policy.

In other words, AI will only become more important in how policymakers see and understand the world. In doing so, it will effectively expand their capacity for processing information, but at the same time it will introduce new uncertainties and complexities into established decision-making protocols.

Predictive roles

Another set of roles for AI might be prediction rather than analysis. In other words, whereas analytical applications of AI are intended to streamline current operations, artificially intelligent systems may offer opportunities for policymakers to understand possible future events.

One such example in the arena of international affairs would be the possibility of modelling complex negotiations. Along with using AI systems to monitor compliance and improve the efficiency of complex international instruments, parties to negotiations (whether economic or strategic in nature) might use sophisticated machine-learning methods to forecast others' positions and tactics.

But a number of moderating factors must be considered. Notably, for instance, while predictive algorithms have been demonstrated with some success in some capacities, they are not yet necessarily more accurate than their human equivalents.[7] Time, accumulated knowledge and increasingly powerful computer hardware may ultimately make them comparably accurate with (or more accurate than) predictions made by humans, but the nature of prediction makes it unlikely that there will be one clear standard of success on this front. Furthermore, as seen in the example of the Tesla fatal accident in 2016, the interface between machine and human understandings of the world creates new potential for miscalculations without necessarily providing

[6] The term 'machine learning' is often (misleadingly) used as a synonym for AI. Machine learning is in fact a specific subset of AI systems whose programming is not wholly written by humans but which 'learns' on the basis of reviewing datasets.
[7] Kanter, J. M. and Veeramachaneni, K. (2015), 'Deep Feature Synthesis: Towards Automating Data Science Endeavors' in *Proceedings of the International Conference on Data Science and Advance Analytics*, IEEE, http://groups.csail.mit.edu/EVO-DesignOpt/groupWebSite/uploads/Site/DSAA_DSM_2015.pdf.

a compensatory benefit. That potential is amplified by the fact that complex negotiations are by definition multi-party: the machine system and human operator are not simply trying to obey a set of rules more efficiently; they are trying to predict the actions of, and overmatch, one or more opponents (who may themselves also be using predictive algorithms).

There may well be benefits arising from machines operating in such predictive capacities – not least the fact that the ability of an artificial system to store and compare numerous historical data points will almost certainly exceed that of a human or group of humans. But, as Heather Roff observes in Chapter 3, for those benefits to be shared equitably, the underlying technology must be accessible – and, by the same token, stringent measures must be taken to protect personalized data that feed into the algorithms. Striking the right balance will be difficult indeed.

It is worth noting, moreover, that AI might take on other predictive roles with a bearing on geopolitics, contributing for instance to more accurate forecasting of elections, economic performance and other relevant events. But such areas are functions less of machine learning and more of the quantity of data available, and so should instead be considered chiefly in that light.

Operational roles

The final category is somewhat different, covering autonomous systems in the more traditional sense of robots. The implications of these applications are likely to be diffuse and indirect, but their potential significance warrants consideration alongside analytical and operational functions.

Autonomous logistical systems are likely to have significant indirect implications for international politics. The day-to-day functioning of the international system would not be expected to change appreciably if truck drivers, ship crews or pilots were replaced with automatons, but the large-scale replacement of existing human labour in these capacities is likely to cause widespread economic and political disruption in the short to medium term, a prospect that Kenn Cukier addresses in Chapter 4.

Autonomous weapons are another significant issue, albeit, as Missy Cummings argues in Chapter 2, perhaps further from wide acceptance than is generally recognized. There is a considerable public debate about the ethics, morality and legality of the development and use of such weapons. The central tenet of this debate is autonomy, which differentiates it from that concerning unmanned drones; the latter are remotely piloted rather than fully autonomous systems.

Ethical questions aside, autonomous weapons do not in and of themselves necessarily change the balance of power.[8] An autonomous strike aircraft makes certain trade-offs relative to a piloted one – essentially greater endurance and expendability as against flexibility and versatility – but it is not inherently a more powerful, game-changing weapon. In the long term, the ability of autonomous systems to react with greater speed than humans are able to might make a difference in some environments (in space,

[8] Scharre, P. (2017), 'A security perspective: Security concerns and possible arms control approaches', in United Nations Office for Disarmament Affairs (UNODA) (2017), *Perspectives on Lethal Autonomous Weapons Systems*, UNODA Occasional Papers – No. 30, November 2017, New York: United Nations, https://www.un.org/disarmament/publications/occasionalpapers/unoda-occasional-papers-no-30-november-2017/, pp. 19–33.

in cyberspace or with respect to hypersonic weapons[9]). Autonomy might also enable the development of new classes of weaponry – swarms of small, interlinked robotic vehicles could create wholly new paradigms of military capability, for example – but given both the logistical and ethical questions arising from such potential systems it is premature to declare the technology ready for deployment, let alone game-changing. One possible exception could be cyberweapons imbued with autonomy and machine learning, with the potential to make them far more effective and adaptable than their existing counterparts.

Even if a fundamental change in warfare driven by autonomy is not an immediate prospect, the consequences in terms of changing norms and standards of how policymakers see and respond to threats are likely to undergo a fundamental change

But, as with civilian autonomous vehicles, the real impact is likely to be indirect for some time to come. A military with autonomous systems may be only marginally more capable than one without – at least in the near term. But even if a fundamental change in warfare driven by autonomy is not an immediate prospect, the consequences in terms of changing norms and standards of how policymakers see and respond to threats are likely to undergo a fundamental change.

How deployment may proceed

To be adopted in any of these three sets of roles, AI will have to demonstrate comparable or greater effectiveness than humans at a comparable cost. If human policymakers are already confident in their human analysts' ability to operationalize a strategically important arms control arrangement, they are unlikely to turn these processes over to a wholly new and unproven system.

Deployment in any case will not simply be a case of handing over the keys or flipping a switch. There will be no 'artificial analysts' ready to simply take on human roles. Rather, AI will increasingly and incrementally be paired with human analysts to take on these tasks. These human–machine 'centaurs' offer the most promise in the foreseeable future, and represent the best chance for a measured transition away from sole human oversight. 'Centaur' pairings theoretically allow the combination of the best qualities of human and machine intelligence: the machine can process enormous quantities of data quickly while the human can spot-check and correct where necessary, as well as understanding, framing and responding to the results in ways that interface with existing policy mechanisms.

Some efforts have already been made to begin to devise an ethical or legal framework for autonomous systems. Most recently, in May 2018 a new declaration on protecting the rights to equality and non-discrimination in machine learning systems was opened for signature by a group of international NGOs. Building on the framework of international human rights law, and emphasizing the responsibilities of public- and private-sector actors, the 'Toronto Declaration' aims to ensure that new machine learning technologies – and AI and related data systems more broadly – incorporate principles of respect for inclusivity and non-discrimination.[10] Meanwhile, the UN has begun to convene working

[9] A class of projectile that travels at greater than five times the speed of sound, considerably faster than existing bullets, rockets or missiles. At such speeds, human reaction time might not be adequate for successful engagement or defence, implying a need for greater autonomy.
[10] Access Now (2018), 'The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems', https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/.

groups under the aegis of the Convention on Certain Conventional Weapons (CCW) to define the legality of the development and use of autonomous weaponry.[11] These steps are welcome, but nascent.

This study does not attempt to cover the entire scope or breadth of the implications of AI technologies for international politics. There are significant areas – among them medicine, public health and law – where AI systems may be transformative in ways that directly impact the processes of the international system in the next decade or two. Nor are the various ways in which AI might intersect – or interfere – with democratic processes considered here – a topic that certainly warrants specific in-depth consideration. Rather, in focusing on the application of AI technologies from military, human security and economic perspectives, the report presents a snapshot of some of the nearer-term consequences of what might be one of the defining trends of the coming decades.

[11] United Nations Office at Geneva (2018), '2018 Group of Governmental Experts on Lethal Autonomous Weapons Systems (LAWS)', https://www.unog.ch/80256EE600585943/(httpPages)/7C335E71DFCB29D1C1258243003E8724?OpenDocument.

2. Artificial Intelligence and the Future of Warfare

M. L. Cummings

In anticipation of this report, the text of this chapter was first published by Chatham House in January 2017.[12]

Introduction

The rise in the use of unmanned aerial vehicles (UAVs) – commonly known as drones – in both military and commercial settings has been accompanied by a heated debate as to whether there should be an outright ban on what some label 'killer robots'.[13] Such robots, which could be in the air, on the ground, or in and under water, theoretically incorporate AI that would make them capable of executing missions on their own. The debate, which has many dimensions and stakeholders, concerns whether artificially intelligent machines should be allowed to execute such military missions, especially if there is a possibility that any human life could be at stake.

Given the complexity of the matter, a working definition of AI is needed. There is no one commonly agreed definition, even among computer scientists and engineers, but a general definition of AI is the capability of a computer system to perform tasks that normally require human intelligence, such as visual perception, speech recognition and decision-making. This definition is, however, inherently oversimplified, since what constitutes intelligent behaviour is also open to debate. Arguably, by this definition a house thermostat is intelligent because it can perceive and adjust the temperature. This is substantially different from AI whereby a UAV selects and engages targets without meaningful human control, which is the common assumption for autonomous weapons.

With the rapidly expanding commercial market for both air and ground autonomous systems, there is evidence of some shifting in AI expertise from military to commercial enterprises

Another critical factor to consider in the debate over autonomous weapons is the increasing inability to disambiguate commercial drone autonomy from that of military UAVs. Indeed, with the rapidly expanding commercial market for both air and ground autonomous systems, there is evidence of some shifting in AI expertise from military to commercial enterprises. As a result, banning an autonomous technology for military use may not be practical given that derivative or superior technologies could well be available in the commercial sector. In addition, the asymmetrical development of the commercial autonomous system market would be likely to result in a lack of expertise for governments and militaries, which could lead to compromised and unsafe autonomous systems, both fully and semi-autonomous.

This chapter presents first a framework explaining the current state of the art for AI, the strengths and weaknesses of AI, and what the future likely holds. Given that the advancement of AI is inextricably linked to the expertise of the engineers developing these systems, the case is then made that the shift in expertise from the military to the commercial sector will further complicate policy discussions on autonomous weapons, and will make it difficult for governments to deploy and manage these systems.

[12] Cummings, M. L. (2017), *Artificial Intelligence and the Future of Warfare*, Research Paper, London: Royal Institute of International Affairs.
[13] Human Rights Watch (2013), 'Arms: New Campaign to Stop Killer Robots', https://www.hrw.org/news/2013/04/23/arms-new-campaign-stop-killer-robots; Human Rights Watch and International Human Rights Clinic (2016), 'Killer Robots and the Concept of Meaningful Human Control: Memorandum to Convention on Conventional Weapons (CCW) Delegates', https://www.hrw.org/news/2016/04/11/killer-robots-and-concept-meaningful-human-control; Future of Life Institute (2015), 'Autonomous Weapons: An Open Letter from AI & Robotics Researchers', http://futureoflife.org/open-letter-autonomous-weapons.

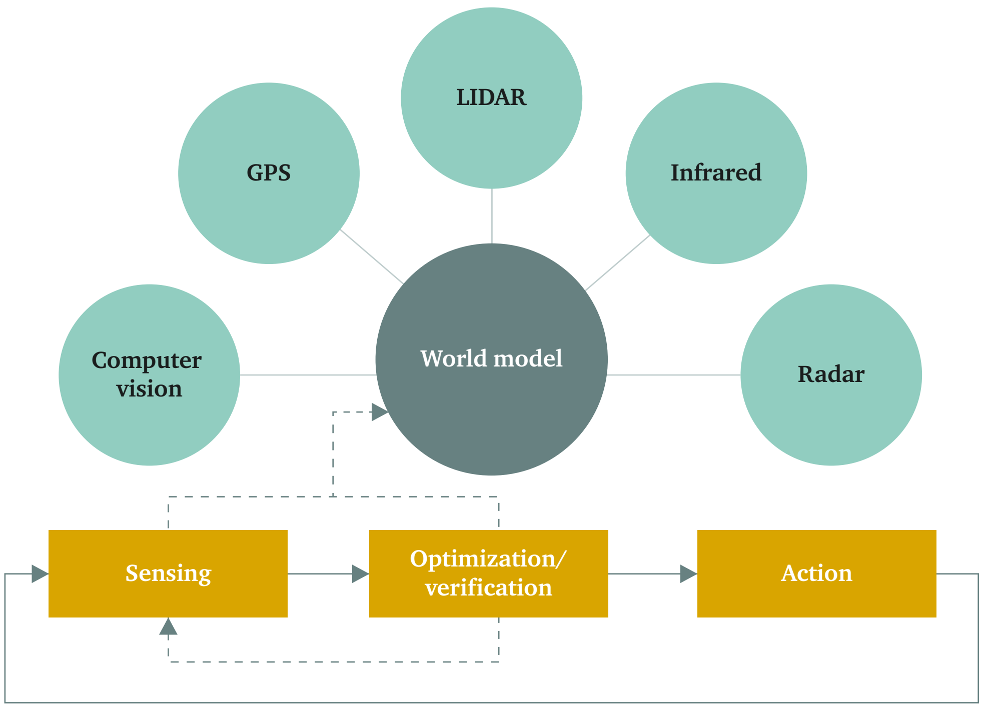
How robots think

An autonomous system is one that reasons probabilistically given a set of inputs, meaning that it makes guesses about best possible courses of action given sensor data input

To better understand the nuances of AI, it is important first to understand the difference between an automated and an autonomous system. An automated system is one in which a computer reasons by a clear if–then–else, rule-based structure, and does so deterministically, meaning that for each input the system output will always be the same (except if something fails). An autonomous system is one that reasons probabilistically given a set of inputs, meaning that it makes guesses about best possible courses of action given sensor data input. Unlike automated systems, when given the same input autonomous systems will not necessarily produce the exact same behaviour every time; rather, such systems will produce a range of behaviours.

Human intelligence generally follows a sequence known as the perception–cognition–action information processing loop, in that individuals perceive something in the world around them, think about what to do, and then, once they have weighed up the options, make a decision to act. AI is programmed to do something similar, in that a computer senses the world around it, and then processes the incoming information through optimization and verification algorithms, with a choice of action made in a fashion similar to that of humans. Figure 1 illustrates how an autonomous system embedded with AI 'thinks' and makes decisions in this way.

Figure 1: How AI of an autonomous system works



Source: Adapted from Hutchins, Cummings, Draper and Hughes (2015).[14]

[14] Hutchins, A. R., Cummings, M. L., Draper, M. and Hughes, T. (2015), 'Representing Autonomous Systems' Self-Confidence through Competency Boundaries', paper presented at the 59th Annual Meeting of the Human Factors and Ergonomics Society, Los Angeles, CA, 26–30 October 2015.

While there are many parallels between human intelligence and AI, there are stark differences too. Every autonomous system that interacts in a dynamic environment must construct a world model and continually update that model (as shown in Figure 1). This means that the world must be perceived (or sensed through cameras, microphones and/or tactile sensors) and then reconstructed in such a way that the computer 'brain' has an effective and updated model of the world it is in before it can make decisions. The fidelity of the world model and the timeliness of its updates are the keys to an effective autonomous system.

Autonomous UAV navigation, for example, is relatively straightforward, since the world model according to which it operates consists simply of maps that indicate preferred routes, height obstacles and no-fly zones. Radars augment this model in real time by indicating which altitudes are clear of obstacles. GPS coordinates convey to the UAV where it needs to go, with the overarching goal of the GPS coordinate plan being not to take the aircraft into a no-fly zone or cause it to collide with an obstacle.

In comparison, navigation for driverless cars is much more difficult. Cars not only need similar mapping abilities, but they must also understand where all nearby vehicles, pedestrians and cyclists are, and where all these are going in the next few seconds. Driverless cars (and some drones) do this through a combination of sensors like LIDAR (Light Detection And Ranging), traditional radars, and stereoscopic computer vision (see Figure 1). Thus the world model of a driverless car is much more advanced than that of a typical UAV, reflecting the complexity of the operating environment. A driverless car computer is required to track all the dynamics of all nearby vehicles and obstacles, constantly compute all possible points of intersection, and then estimate how it thinks traffic is going to behave in order to make a decision to act.

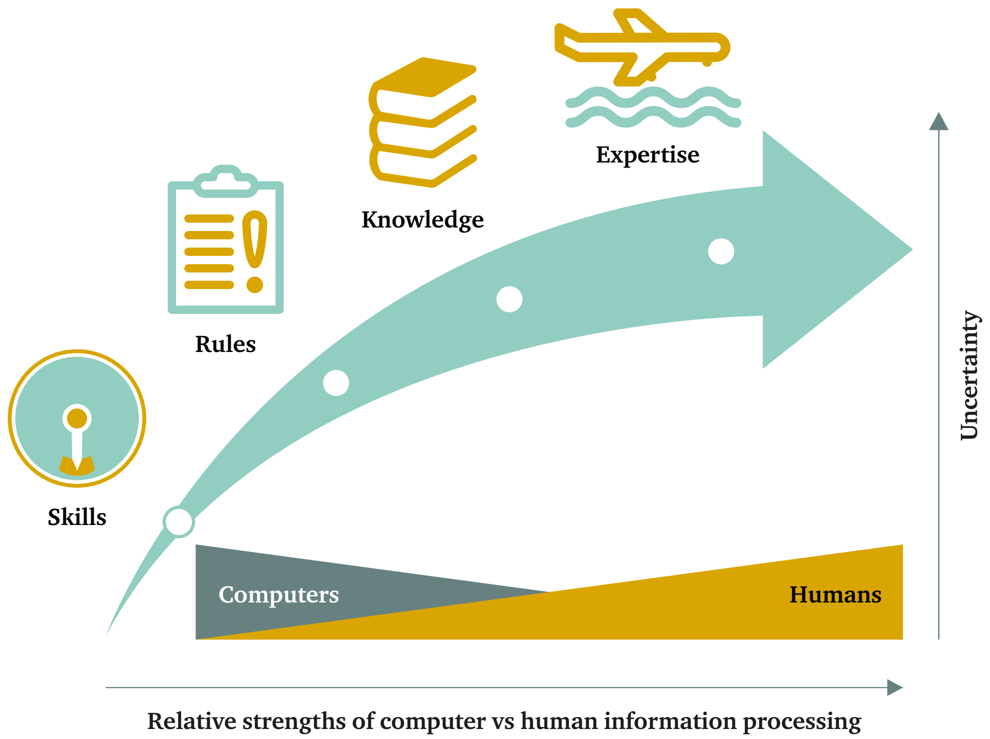
Indeed, this form of estimating or guessing what other drivers will do is a key component of how humans drive, but humans do this with little cognitive effort. It takes a computer significant computation power to keep track of all these variables while also trying to maintain and update its current world model. Given this immense problem of computation, in order to maintain safe execution times for action a driverless car will make best guesses based on probabilistic distributions. In effect, therefore, the car is guessing which path or action is best, given some sort of confidence interval. The best operating conditions for autonomous systems are those that promote a high-fidelity world model with low environment uncertainty, a concept that will be further discussed in the next section.

Balancing tasks between humans and robots

Given this understanding of the basics of autonomous robot reasoning, how should we think about the design of autonomous systems, particularly in terms of the extent to which humans should be involved? It is important first to understand when systems can and should be supervised by humans. This is a decision involving clear technical questions (such as whether a computer vision system can generate an image of sufficient resolution to make an accurate decision) as well as ethical and policy considerations (such as whether a robot should be allowed to take the life of a human being). Detailed understanding of the construction and capabilities of such military AI systems is needed in order to form cogent arguments around this polarizing issue.

Figure 2 depicts the stages of reasoning that any agent must possess in order to deal with increasingly complex decision-making scenarios, including those of autonomous weapons operation. This is an extension of Rasmussen's SRK (skills, rules and knowledge-based behaviours) taxonomy, modified to explicitly represent expertise and uncertainty.[15]

Figure 2: The relationship between uncertainty and skill, rule, knowledge and expert reasoning



Source: Adapted from Cummings (2014).[16]

Skills-based behaviours are sensory-motor actions that for humans become highly automatic after some period of training. Such skills rely on a tight coupling of the perception–cognition–action loop, which effectively means that actions must typically come within seconds of a stimulus. An example of a skills-based activity is flying an aircraft. In training, human pilots spend the bulk of their time learning to interpret dials and gauges and thereby adjust the aircraft controls appropriately quickly to make sure the actual state of the aircraft matches the intended state. Automated agents do this as well, through an equations-based feedback loop that relies on the quality of the input data from the sensors that tell the computer where the aircraft is and where it should be.

As the cognitive continuum increases in complexity (depicted by the arrow in Figure 2), the need for rule-based behaviours arises. Rule-based behaviours are those actions that can be represented by subroutines, stored rules or procedures. Checklists are common cognitive-aiding tools for humans executing rules. To continue with the example

[15] Rasmussen, J. (1983), 'Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models', *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3), pp. 257–266.
[16] Cummings, M. L. (2014), 'Man vs. Machine or Man + Machine?' *IEEE Intelligent Systems*, 29(5), pp. 62–69.

of aviation, pilots rely significantly on procedures to help them manage the complexity of various tasks. For instance, when a fire-light illuminates or another subsystem indicates a problem, pilots are trained to first stabilize the aircraft (a skill) but then turn to the manual to determine the correct procedure (rule following). Such codified procedures are necessary since there are far too many solutions to possible problems to be remembered. Some interpretation of procedures is required in many scenarios, especially as uncertainty and complexity increases, and this is particularly common in cases of multiple and compound problems.

Knowledge-based reasoning occurs when an established set of rules does not necessarily match the current situation, and fast mental simulations could be needed for resolution. Mental models – cognitive representations of the external world – are built over time and assist in the formulation and selection of plans, particularly in conditions of uncertainty. The so-called 'miracle on the Hudson', the landing of US Airways Flight 1549 on the Hudson River in 2009, is an example of a knowledge-based behaviour depicted in Figure 2: during an emergency, the flight captain had to decide whether to ditch the aircraft or attempt to land it at a nearby airport. This is the quintessential knowledge-based scenario, reflecting a high degree of uncertainty that required the captain to develop a mental model of the environment and the state of the aircraft. The resultant fast mental simulation meant that he chose the ditching option, with clear success.[17] At the time, no autopilot system had the capability to respond in such a manner (as remains the case, although there is active research in this area).

In humans, the ability to cope with the highest situations of uncertainty is one of the hallmarks of a true expert, but in comparison such behaviour is very difficult for computers to replicate

Expert behaviours sit at the top of the reasoning behaviours, which build on knowledge-based reasoning. Expertise leverages judgment and intuition as well as the quick assessment of a situation, especially in a time-critical environment such as weapons release. Experts typically make difficult decisions in a fast and frugal manner, since comparing all possible plan alternatives is time-intensive – particularly in the face of uncertainty.[18] In humans, the ability to cope with the highest situations of uncertainty is one of the hallmarks of a true expert, but in comparison such behaviour is very difficult for computers to replicate. As depicted in Figure 2, skill-based tasks are easiest to automate, since they are by definition highly repetitive with inherent feedback loops that can be controlled through mathematical representations. A critical assumption, nonetheless, is that the appropriate sensors are in place. Given the if–then–else structure of rule-based behaviours, these are also potentially good candidates for automation. However, as shown in Figure 2, as uncertainty increases, rule-based reasoning gives way to knowledge-based reasoning, requiring effective management of uncertainty and true expertise.

Navigation, for example, is very rule-based, in that given a goal state (the destination), the best path can be objectively defined given the rules about traffic flow and knowledge of vehicle dynamics. All the same, uncertainty in such a domain can make the introduction of autonomy difficult. As already noted, navigation for drones is relatively straightforward, since mature sensor capabilities and a low-density obstacle environment mean that uncertainty is low. In the car-driving scenario, by contrast, uncertainty is much higher because the sensors are not as reliable as they are in aircraft, and the potential obstacle field is significantly more dense.

---

[17] In 2009 US Airways Flight 1549 suffered a dual engine failure after take-off from New York's LaGuardia airport due to ingestion of birds in both engines. Unable to reach any prepared landing strip for an emergency landing, the pilot, Capt. Chesley Sullenberger, elected to ditch the aircraft into the Hudson River. There were no fatalities, and all 155 passengers and crew were able to evacuate the aircraft; all were subsequently rescued by boat.
[18] Gigerenzer, G., Todd, P. M. and ABC Research Group (1999), *Simple Heuristics That Make Us Smart*, Oxford and New York: Oxford University Press.

It is at the rule-based level of reasoning where the shift between needing automated versus autonomous behaviours starts to become evident. Some higher-level reasoning begins here, but the uncertainty also starts to grow as well – especially in the presence of an incomplete rule set. For instance, the Global Hawk military UAV works at a rule-based level whereby it is able to land itself if it loses communication, but it has not yet been demonstrated that such an aircraft can reason under all the situations it might encounter. The latter would require a higher level of reasoning.

Rule-based reasoning may assist decision-makers (human and/or computer) in determining possible courses of action. Where there is high uncertainty, however, it is often difficult to understand which set of rules applies

Knowledge-based behaviours and resulting expertise represent the most advanced forms of cognitive reasoning that typically occur in domains where uncertainty is the highest, as shown in Figure 2. Rule-based reasoning may assist decision-makers (human and/or computer) in determining possible courses of action. Where there is high uncertainty, however, it is often difficult to understand which set of rules applies. It is in these uncertain scenarios – which are by definition vague and ambiguous – that algorithms may not be able to understand the solution space, much less achieve a feasible solution.

The key question for any autonomous system engaged in a safety-critical task (such as weapons release) is whether that system can resolve ambiguity in order to achieve acceptable outcomes. It is conceivable that an autonomous drone could be given a mission to hit a static target on a military installation with a high probability of success. Indeed, many countries have missiles that can do just that. However, could an autonomous drone targeting an individual understand from its real-time imagery that a specific person has been found and that releasing a weapon will kill only that person and no innocent bystanders? Currently, the answer to this question is no.

The power of human induction – i.e. the ability to form general rules from specific pieces of information – is critical in a situation that requires both visual and moral judgment and reasoning. For humans, induction that drives such judgments is necessary to combat uncertainty. Computer algorithms – especially those that are data-driven like typical algorithms that fall under the category of AI – are inherently brittle, which means that such algorithms cannot generalize and can only consider the quantifiable variables identified early on in the design stages when the algorithms are originally coded.[19] Replicating the intangible concept of intuition, knowledge-based reasoning and true expertise is, for now, beyond the realm of computers. There is significant research currently under way to change this, particularly in the machine learning/ AI community, but progress is slow.

IBM's Watson, composed of 90 servers, each with a 3.5 GHz core processor, is widely referenced as a computer that has the capacity to outmatch human reasoning abilities.[20] As such, it is often popularly stated to be an 'intelligent' computer. However, people confuse the ability of a computer – which has been tuned by humans for a highly specific task of searching vast databases and generating formulaic responses – with an entity that exhibits knowledge. Watson leverages AI through natural language processing and pattern matching, but it is operating in environments in which uncertainty is low.

[19] Smith, P. J., McCoy, C. E. and Layton, C. (1997), 'Brittleness in the design of cooperative problem-solving systems: The effects on user performance', *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 27(3), pp. 360–371.
[20] Deedrick, T. (2011), 'It's Technical, Dear Watson', *IBM Systems Magazine*, http://www.ibmsystemsmag.com/ibmi/trends/whatsnew/It%E2%80%99s-Technical,-Dear-Watson.

Some claim that machine learning and deep learning[21] approximate human intelligence, but at present these tools basically detect patterns that are significantly tuned by humans and must be interpreted by humans to be useful. As a result, the advances that they represent are evolutionary and not revolutionary. One critical limitation of machine learning is that, as a data-driven approach, it fundamentally relies on the quality of the underlying data and thus can be very brittle. There are no guarantees that a computer leveraging machine learning algorithms can detect a pattern or event never previously encountered, or even scenarios that are only slightly different. As uncertainty grows, therefore, these tools become less useful.

For example, a highly publicized machine learning algorithm has been able to identify objects (e.g. dog, cat, truck) in images with 15.8 per cent accuracy in a world containing some 22,000 object categories.[22] When that world is collapsed into 1,000 categories of objects, other algorithms can achieve up to 60–70 per cent accuracy.[23] In both of these studies, 10 million labelled images were required to 'train' the algorithms. In comparison, humans need far fewer examples to 'learn' from, and have the ability to distinguish and accurately name far more than 22,000 individual objects.

So while computer vision algorithms may be able to distinguish a car from a bicycle at close range for the purposes of driverless car navigation, such algorithms are far from perfect – as demonstrated in March 2018 by the death of a pedestrian in Tempe, Arizona, when a self-driving car failed to detect her in reportedly near-ideal conditions.[24] Moreover, it is multiple orders of magnitude more difficult for a computer to identify a specific truck carrying specific people with enough certainty to launch a weapon. So while UAVs are excellent platforms for obtaining an image, letting the UAV or any autonomous system decide whether a specific target meets the criteria for weapons release is still several years away.

While the consequences of an autonomous central heating thermostat wrongly guessing a homeowner's arrival time are relatively trivial, the same cannot be said of an autonomous weapon making an inaccurate guess in its particular operating environment

The inability of computers to identify specific targets with a high degree of certainty highlights the probabilistic nature of knowledge-based reasoning. Whether humans or computers do it, both are guessing with incomplete information based on prior probabilities about an outcome. While the consequences of an autonomous central heating thermostat wrongly guessing a homeowner's arrival time are relatively trivial, the same cannot be said of an autonomous weapon making an inaccurate guess in its particular operating environment.

The big picture

The future of AI in military systems is directly tied to the ability of engineers to design autonomous systems that demonstrate independent capacity for knowledge- and expert-based reasoning as illustrated in Figure 2. There are no such autonomous systems currently in operation. Most ground robots are teleoperated, essentially meaning that a human is still directly controlling a robot from some distance away as though via a virtual

---

[21] Deep learning is a branch of machine learning that attempts to match the processes by which humans learn using especially large and complex artificial 'neural networks'. See Hof, R. (2013), 'Deep learning: With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart', in *MIT Technology Review*, https://www.technologyreview.com/s/513696/deep-learning/.
[22] Le, Q. V. et al. (2012).
[23] Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y. (2014), 'OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks', paper presented at the International Conference on Learning Representations, Banff, AB, 14–16 April 2016.
[24] Griggs, T. and Wakabayashi, D. (2018), 'How a Self-Driving Uber Killed a Pedestrian in Arizona', *New York Times*, 21 March 2018, https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html.

extension cord. Most military UAVs are only slightly more sophisticated: they have some low-level autonomy that allows them to navigate, and in some cases land, without human intervention, but almost all require significant human intervention to execute their missions. Even those that take off, fly over a target to capture images, and then return home still operate at an automated and not autonomous level, and do not reason on the fly as true autonomous systems would.

While current operational systems are more automatic than autonomous, there are significant global efforts in the research and development (R&D) of autonomous systems. Incremental progress in such military system development is occurring in many countries in air, ground, on-water and underwater vehicles with varying degrees of success. Several types of autonomous helicopter that can be directed with a smartphone by a soldier in the field are in development in the US, in Europe and in China. Autonomous ground vehicles such as tanks and transport vehicles are in development worldwide, as are autonomous underwater vehicles. In almost all cases, however, the agencies developing these technologies are struggling to make the leap from development to operational implementation.

There are many reasons for the lack of success in bringing these technologies to maturity, including cost and unforeseen technical issues, but equally problematic are organizational and cultural barriers. The US has, for instance, struggled to bring autonomous UAVs to operational status, primarily as a result of organizational in-fighting and prioritization in favour of manned aircraft.[25] For example, despite the fact that the F-22 aircraft has experienced significant technical problems and has flown little in combat, the US Air Force is considering restarting the F-22 production line – in itself an extremely costly option – as opposed to investing in more drone acquisitions. Beyond the production line, moreover, the hourly operational cost of the F-22 is $68,362, as compared with the Predator's $3,679;[26] the latter can perform most of the same core functions of an F-22 save for air-to-air combat missions, which the F-22 itself could not previously perform due to technical problems.

To give another example, the US Navy's X-47 was intended to be developed as an autonomous fighter/bomber aircraft, but despite many successful sea trials it is now slated to operate as a tanker for aerial refuelling – a far cry from its original (achievable) purpose. Both the US Air Force and Navy have chosen to use the vast majority of aircraft acquisition funds for the manned F-35 Joint Strike Fighter, even though the programme has been beset with management and engineering problems, and whose relevance is contested particularly in light of advancing autonomous technologies.[27] For many in the military, UAVs are acceptable only in a support role, but they threaten the status quo if allowed to take the most prestigious, 'tip-of-the-spear' jobs.

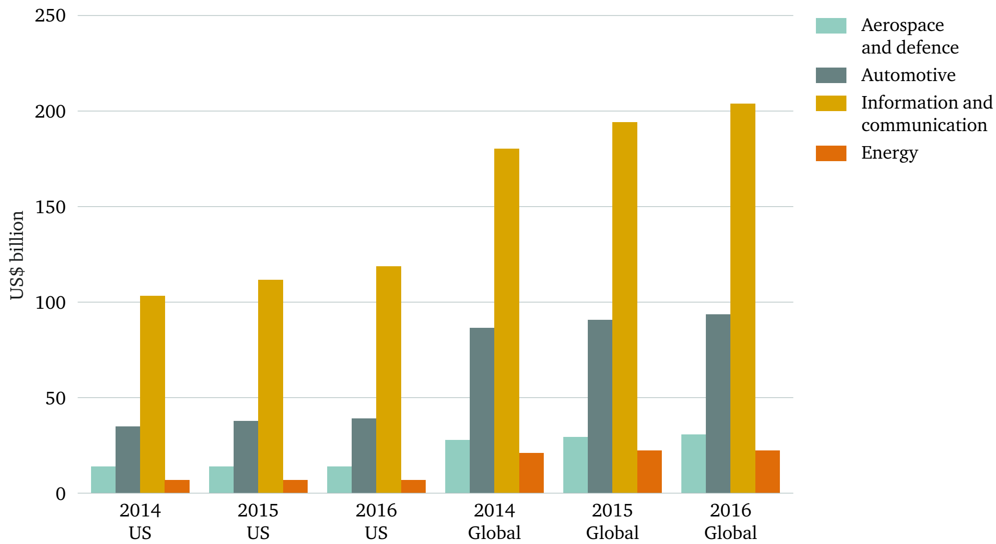
There are, however, other organizational issues limiting the operational implementation of autonomous systems, and one that is increasingly problematic is the shift in advanced development from military to commercial settings. A metaphorical arms race is in progress in the commercial sphere of autonomous systems development. Military autonomous systems development has been slow and incremental at best, and pales in comparison with the advances made in commercial autonomous systems such as drones, and especially in driverless cars.

[25] Spinetta, L. and Cummings, M. L. (2012), 'Unloved Aerial Vehicles: Gutting its UAV plan, the Air Force sets a course for irrelevance', *Armed Forces Journal*, November 2012, pp. 8–12.
[26] Thompson, M. (2013), Costly Flight Hours, *Time*, 2 April 2013.
[27] Hendrix, J. (2015), *Retreat from Range: The Rise and Fall of Carrier Aviation*, Washington, DC: Center for a New American Security.

Driverless car development originated with a Defense Advanced Research Projects Agency (DARPA) programme in 2004. When the programme ended in 2007, driverless cars could move only slowly through closed courses, and not without accidents. A decade later, industry is on the verge of commercializing driverless cars around the world. This rapid progress is a result of the significant industry-sponsored R&D investment, as well as competition for the multi-billion-dollar automotive consumer market. Meanwhile – and paradoxically, given the origins of the technology – there has been very little progress in military autonomous vehicle development.

The inability of the military to advance its autonomy programmes, not only on the ground but also in the air and in other domains, is evidently linked to the growth in autonomous systems in the commercial market, particularly driverless cars. Figure 3 shows R&D spending in the three-year period 2014–16 in three key sectors: aerospace and defence; automotive; and information and communication. These sectors are core to the development of autonomous systems, and so tracking spending there gives insight into the speed and extent of innovation.

Figure 3: R&D spending by sector, 2014–16 (US$ billion)



Source: Industrial Research Institute (2016).[28]

Only a very small proportion of defence R&D money is invested in autonomous military systems

The aerospace and defence sector is responsible for the bulk of the development of military autonomous systems. However, as shown in Figure 3, R&D spending is far below that of the other two sectors (only around 15 per cent of the global information and communication sector, for example). In the US, moreover, spending has actually been declining. Autonomous systems development is not a priority for the US defence industry, and competes for investment with development of traditional platforms like narrowly capable manned fighter aircraft (e.g. the F-35) and extremely costly laser weapons. The outcome is that only a very small proportion of defence R&D money is invested in autonomous military systems.

[28] Industrial Research Institute (2016), '2016 Global R&D Funding Forecast', supplement to *R&D Magazine*, Winter 2016.

In contrast, not only is R&D spending in the automotive sector three times that of the aerospace and defence industry, but it is growing in both the US and the global markets. A key factor in spending in this sector is the development of driverless car technology – as reflected in the intensely competitive market for start-ups such as Cruise Automation, bought by General Motors for more than $1 billion in 2016.[29]

The information and communication sector is another critical stakeholder in the development of autonomous systems. Industries within the sector specialize in the development of software, including machine learning and AI that are core capabilities for these advanced systems. To give some examples, X (formerly known as Google X) has both drone and driverless car research programmes; Facebook and Amazon have drone development programmes; and Apple is thought to have an autonomous car development project. Given that the information technology companies are spending far more on R&D than does the aerospace and defence sector (Figure 3), it is clear why there has been far greater progress in commercial autonomous systems development.

As regards the future of warfare as it is linked to AI, the large disparity in commercial versus military autonomous R&D spending could have a cascading effect on the types and quality of autonomy that are eventually incorporated into military systems. One critical issue in this regard is whether defence companies will have the capacity to develop and test safe and controllable autonomous systems, especially those that fire weapons.

Engineers who design such systems must have strong hardware as well as software expertise. However, there are a limited number of universities graduating students in robotics, controls, mechatronics and related fields with the technical prowess for such jobs, and so there is fierce competition for highly qualified personnel. This demand is what drives $1 billion start-up acquisitions, as well as buy-outs of university research groups such as that by Uber, in 2015, of 40 highly skilled roboticists from the National Robotics Engineering Center at Carnegie Mellon University. Boston Dynamics, at one time the leading US military R&D robotics company, was bought by Google in 2013 for domestic robot development, and was reported by a number of sources in 2016 to be a target for acquisition by Toyota for its driverless car development programme.[30] Eventually, in mid-2017 it was announced that a subsidiary of Japan's SoftBank Group had entered into a definitive agreement to acquire Boston Dynamics from Google's parent company, Alphabet.[31]

The global defence industry is falling behind its commercial counterparts in terms of technology innovation, with the gap only widening as the best and brightest engineers move to the commercial sphere

With such a hotly competitive market for roboticists and related engineers across these industries, the aerospace and defence sector, where funding lags behind, is less appealing to the most able personnel. As a result, the global defence industry is falling behind its commercial counterparts in terms of technology innovation, with the gap only widening as the best and brightest engineers move to the commercial sphere. This comparative lack of expertise means that military autonomous systems eventually fielded could be deficient, or lacking in appropriate safeguards and testing. So while the debate over whether autonomous weapons should be banned is clearly an important one, a more immediate issue is the ability of defence industries to field safe semi-autonomous systems, much less fully autonomous ones.

---

[29] Ibid.
[30] Brian, M. (2016), 'Toyota is the top bidder for robotics pioneer Boston Dynamics', Engadget, 1 June 2016, https://www.engadget.com/2016/06/01/toyota-alphabet-boston-dynamics.
[31] SoftBank Group (2017), 'SoftBank Announces Agreement to Acquire Boston Dynamics', Press Release, 9 June 2017, https://www.softbank.jp/en/corp/news/press/sb/2017/20170609_01/.

While some may say that the current distribution of both R&D and expertise is the inevitable outcome of competition and a free market, this does not fully acknowledge the reality of a fundamental shift in technology prowess whereby militaries will start to significantly lag in autonomous system capabilities as compared with commercial systems. The average American is more likely to have a driverless vehicle before soldiers on the battlefield do, and terrorists will potentially be able to buy drones on the internet with as much or greater capability than those available to the military.

If defence companies and governments continue down a path of relative AI illiteracy, could this enable a potential power shift such that critical AI services will be leased via Google, Amazon or Facebook

This imbalance in technology access will no doubt introduce new unforeseen and disruptive dynamics for military operations. For example, if defence companies and governments continue down a path of relative AI illiteracy, could this enable a potential power shift such that critical AI services will be leased via Google, Amazon or Facebook? If militaries are relegated to buying robots and AI services such as image analysis from commercial, off-the-shelf suppliers, this would undoubtedly affect military readiness in both the short and the long term.

The gap between historical military superiority in UAV development and the present capabilities of the commercial sector is closing, as evidenced by the increasing number of military-grade drones offered for sale via the internet. Footage showing weaponization of drones, once thought to be the exclusive domain of established militaries, can now be regularly be found on YouTube. It is unlikely that small entrepreneurial companies will produce drones that could compete with highly advanced UAVs such as the US Air Force's Predator, but there will certainly be competition from such companies in terms of developments in surveillance and communication. And while the US military's advanced research arm, DARPA, has had difficulty in fielding a drone that can transport troops, China's EHang has purportedly built a commercial drone that will transport passengers, and several commercial companies are developing versions of autonomous flying cars for general public use.[32]

Given the current extent of commercial development of drones and other robotic systems, there are other important considerations such as the possible latent consequences of companies and countries that rush AI technologies to market – as against nation states that tend to take more conservative approaches. Fielding nascent technologies without comprehensive testing could put both military personnel and civilians at undue risk. However, the rapid development of commercial autonomous systems could normalize the acceptance of autonomous systems for the military and the public, and this could encourage state militaries to fund the development of such systems at a level that better matches investment in manned systems. Meanwhile, it remains unclear how the rise of autonomous drones for civilian use could influence popular attitudes and perceptions concerning autonomous military platforms, including weapons.

The thorny path ahead

Although it is not in doubt that AI is going to be part of the future of militaries around the world, the landscape is changing quickly and in potentially disruptive ways. AI is advancing, but given the current struggle to imbue computers with true knowledge and expert-based behaviours, as well as limitations in perception sensors, it will be many years before AI will be able to approximate human intelligence in high-uncertainty settings – as epitomized by the fog of war.

[32] Forum (2016), 'Future of urban mobility: My kind of flyover', Airbus Forum, http://www.airbus.com/newsroom/news/en/2016/12/My-Kind-Of-Flyover.html.

Autonomous guidance systems for missiles on drones will likely be strikingly similar to those that deliver packages, so banning one could affect the other

Given the present inability of AI to reason in such high-stakes settings, it is understandable that many people want to ban autonomous weapons, but the complexity of the field means that prohibition must be carefully scoped. Fundamentally, for instance, does the term autonomous weapon describe the actual weapon – i.e. a missile on a drone – or the drone itself? Autonomous guidance systems for missiles on drones will likely be strikingly similar to those that deliver packages, so banning one could affect the other. And how will technologies be treated that emerge from the growing commercial market, which is expected to leapfrog some aspects of military capability and possibly change public perception?

The impact of the rapid expansion of the commercial market on autonomous systems development cannot be overstated, and an even bigger problem in the short term is how to fully understand the global implications of the discernible shift in the power base of AI expertise from the military to commercial enterprises. Machines, computers and robots are getting 'smarter' primarily because roboticists and related engineers are getting smarter, so this relatively small group of expert humans is becoming a critical commodity. Universities have been slow to respond to this demand, and governments and industry have also lagged behind in providing scholarship mechanisms to incentivize students in the field of AI.

Ultimately, the growth in the commercial information technology and automotive sectors, in terms of both attracting top talent and expanding autonomous systems capabilities in everyday commercial products, could be a double-edged sword that will undoubtedly affect militaries around the world in as yet unimagined ways.

3. Advancing Human Security Through Artificial Intelligence

Heather M. Roff

In anticipation of this report, the text of this chapter was first published by Chatham House in May 2017.[33]

Introduction

Over the next two decades, human security will be confronted by significant challenges. With continuing global warming there will be increased temperatures, rising sea levels and more extreme weather events.[34] These changes will lead to a scarcity of resources, particularly of water, food and energy.[35] The hardest hit areas of the globe are most likely to be those already suffering from various types of instability, violence and unrest, such as sub-Saharan Africa, Pakistan, and parts of the Middle East and North Africa.[36] The confluence of climate and political refugees will undoubtedly compromise local, regional and the international community's ability to secure individuals from fear and want.

With an estimated 50 billion connected devices, all generating mass amounts of data, information will become an even more powerful tool for development, coordination, persuasion and coercion

Concomitantly, growing connectedness via social media and changes in labour and production due to advancing technology proliferation will also place new stresses on the world economy, as well as create new shifts in political and economic power. Microsoft has predicted that by 2025, 4.7 billion people will use the internet – just over half the world's expected population at that time – and, of that number, 75 per cent of users will be in emerging economies.[37] With an estimated 50 billion connected devices, all generating mass amounts of data, information will become an even more powerful tool for development, coordination, persuasion and coercion.[38] Moreover, these individuals will enter new (and old) economic market sectors, and be faced with increasing automation and the stresses of wage devaluation.

In this future world, increasingly divided on demographic, economic and technological lines, achieving human security will not be without its difficulties. Systemic challenges, such as climate change and war, and more localized threats like social, economic or political disruptions are almost certain.

One way to meet these challenges is through novel applications of technology, and of AI in particular. AI holds much promise to enable the international community, governments and civil society to predict and prevent human insecurity. With increased connectivity, more sophisticated sensor data and better algorithms, AI applications may prove beneficial in securing basic needs and alleviating or stopping violent action.

[33] Roff, H. M. (2017), *Advancing Human Security through Artificial Intelligence*, Research Paper, London: Royal Institute of International Affairs.
[34] Global Facility for Disaster Reduction and Recovery (2014), *Understanding Risk in an Evolving World: Emerging Best Practices in Natural Disaster Risk Assessment*, World Bank, https://www.gfdrr.org/sites/gfdrr/files/publication/Understanding_Risk-Web_Version-rev_1.8.0.pdf.
[35] Veldkamp, T. I. E, Wada, Y., de Moel, H., Kummu, M., Eisner, S., Aerts, J. C. J. H., and Ward, P. J. (2015), 'Changing Mechanism of Global Water Scarcity Events: Impacts of Socioeconomic Changes and Inter-annual Hydro-climatic Variability,' *Global Environmental Change*, 32, pp.18–29, http://www.sciencedirect.com/science/article/pii/S0959378015000308.
[36] World Bank (2011), 'Climate Change Adaptation and Natural Disasters Preparedness in the Coastal Cities of North Africa'; Petley, D. N. (2010), 'On the Impact of Climate Change and Population Growth on the Occurrence of Fatal Landslides in South, East and SE Asia', *Quarterly Journal of Engineering Geology & Hydrogeology* 43(4), pp. 487–96, http://qjegh.lyellcollection.org/cgi/doi/10.1144/1470-9236/09-001; Mueller, V., Gray, C., and Kosec, K. (2014), 'Heat Stress Increases Long-term Human Migration in Rural Pakistan', *Nature Climate Change* 4(3), pp. 182–85, https://www.nature.com/articles/nclimate2103.
[37] Burt, D., Kleiner, A., Paul Nicholas, J., and Sullivan, K. (2014), 'Cyberspace 2025: Today's Decisions, Tomorrow's Terrain', Microsoft, https://www.microsoft.com/en-us/cybersecurity/content-hub/cyberspace-2025.
[38] In 2016 internet traffic reached 1.3 Zettabytes. (A Zettabyte is 1,000,000,000,000,000,000,000 bytes of information.) This is a 50 per cent increase from 2011–14. Thus it is likely that if this trend continues, then by 2025 traffic will increase to 4.38 Zettabytes. However, if there are advances in storage, more devices, and more streaming, then traffic will exceed this. See, http://highscalability.com/blog/2012/9/11/how-big-is-a-petabyte-exabyte-zettabyte-or-a-yottabyte.html; http://www.tvtechnology.com/resources/0006/welcome-to-the-zettabyte-era/278852.

This chapter lays out first the principles of the UN approach to human security, as well as more critical viewpoints. Next, it argues that many of the conflict and development problems facing the international community, states and civil society can be ameliorated or solved by advancements in AI. In particular, algorithms adept at planning, learning and adapting in complex data-rich environments could permit stakeholders to predict and coordinate responses to many types of humanitarian and human security related situations. Finally, the case is made that to ensure broad access, transparency and accountability, especially in countries that may be prone to human security emergencies, the relevant AI ought to be open source and sensitive to potential biases.

Human security

Human security is a concept that takes the human – as opposed to the state – as the primary locus of security. As former UN High Commissioner for Refugees Sadako Ogata has written, 'Traditionally, security issues were examined in the context of "State security", i.e. protection of the State, its boundaries, its people, institutions and values from external attacks. People were considered to be assured of their security through protection extended by the State.'[39] Yet, with changes in the post-Cold War era, where external threats to state security declined and internal threats of intra-state violence increased, many policymakers, practitioners and scholars required a new lens through which to understand these internal conflicts.

Indeed, in 1994, the UN *Human Development Report* asserted:

> [W]ithout the promotion of people-centred development, none of our key objectives can be met – not peace, not human rights, not environmental protection, not reduced population growth, not social integration. It will be a time for all nations to recognize that it is far cheaper and far more human to act early and to act upstream than to pick up the pieces downstream, to address the root causes of human insecurity rather than its tragic consequences.[40]

From 1994 onwards, many different avenues for examining the concept of human security emerged.[41] Central to all, however, was the focus on the nexus between development, human rights (protection and promotion), and peace and security. The premise that people possess dignity logically entailed that they ought to be 'free from fear' and 'free from want'.[42] To establish what this expansive formulation meant, the 1994 *Human Development Report* identified seven elements comprising human security.[43]

---

[39] Ogata, S. (2015). 'Striving for Human Security', *United Nations Chronicle*, No.1 and 2, p. 26, https://unchronicle.un.org/article/striving-human-security.
[40] United Nations Development Programme (1994), *Human Development Report*, New York and Oxford: Oxford University Press, p. iii.
[41] See for example Human Security Now (2003), the United Nations Trust Fund for Human Security, 2005 World Summit Outcome Document, especially paragraph 143. General Assembly Resolution 66/290 of 2012 addressing explicitly human security and requesting additional reports on lessons learned from 'human security experiences at the international, regional and national levels', A/Res/66/290.
[42] Ibid.
[43] 1994 *Human Development Report* findings, as cited in: Paris, R. (2001), 'Human Security: Paradigm Shift or Hot Air?', *International Security*, 26(2), p. 90.

Table 1: Human security dimensions

Object of security	Content
Economic	Freedom from poverty
Food	Access to food
Health	Access to healthcare and protection from disease
Environmental	Protection from environmental pollution and depletion
Personal	Physical safety (e.g. freedom from torture, war, criminal attacks, domestic violence, drug use, suicide and traffic accidents)
Community	Survival of traditional cultures, ethnic groups and the physical security thereof
Political	Freedom to enjoy civil and political rights, freedom from political oppression

Source: 1994 *Human Development Report* findings, as cited in: Paris (2001), 'Human Security: Paradigm Shift or Hot Air?'.

Human security should be seen as complementary to state security, and measures taken to uphold human rights and build local or regional security capacities through non-coercive measures will simultaneously generate greater stability and development

Following from this, the UN also framed human security as emerging from the achievement of 'sustainable development' and various established international development goals.[44] Human security should be seen as complementary to state security, and measures taken to uphold human rights and build local or regional security capacities through non-coercive measures will simultaneously generate greater stability and development.

However, despite the UN rhetoric, the notion is not without critics. Some claim that it is 'so broad that it is difficult to determine what, if anything, might be excluded from the definition of human security'.[45] The problem, of course, is that if human security as a concept includes such extensive facets of human existence, in reality it means little and impedes the formulation of sound policy. Others point out that the two key elements that define human security have not been treated equally, with progress on the 'freedom from want' portion subjugated to issues related to war and violence, in an attempt to make 'freedom from fear' a reality.[46] Such prioritization reflects various realities of power politics, and demonstrates how some states view their obligations towards capacity-building in areas that have little, if any, strategic or economic interest for them. Indeed, even responses to global health crises appear to mirror power politics and national security interests.[47]

From a practical perspective, difficulties of adequately and appropriately responding to potential, emerging or ongoing human security crises are endemic. One might claim this is due to the fact that the concept is over expansive, but this objection notwithstanding, achieving human security may have more to do with the inability of various stakeholders, such as the UN, civil society and nation states, to monitor, predict and react to a crisis. Since there are linkages between development, human rights and security, the number of different actors with varying priorities and knowledge bases is high. These actors become disconnected, and may even be forced to work against one another for resources or funding. Lack of communication and information exchange between these actors may only exacerbate problems.

[44] A/RES/66/290.
[45] Paris (2001), 'Human Security: Paradigm Shift or Hot Air?'.
[46] Schittecatte, C. (2006), 'Toward a More Inclusive Global Governance and Enhanced Human Security' in Maclean, S., Black, D.R., and Shaw, T. M. (eds) (2006), *A Decade of Human Security: Global Governance and New Multilateralisms*, Ashgate Press, p. 132.
[47] O'Manique, C. (2006), 'The 'Securitization' of HIV/AIDS in Sub-Saharan Africa: A Critical Feminist Lens' in Maclean, Black and Shaw (eds) (2006), *A Decade of Human Security: Global Governance and New Multilateralisms*, p. 165.

Thus, to counter some of these objections, especially in light of the challenges in the coming 10–15 years, it is necessary to devise novel approaches to ameliorate human insecurity and vulnerability. Specifically, by taking a closer look at how new AI applications can help a variety of stakeholders predict, plan and respond to human security crises.

Securing the human through AI

The expansive and interconnected set of factors that affect human security is not the only challenge to alleviating human insecurities.[48] There are three antecedent constraints on human security-related activities: the inability to know about threats in advance; the inability to plan appropriate courses of action to meet these threats; and the lack of capacity to empower stakeholders to effectively respond. Tackling these constraints could save thousands of lives. The use of AI is one potential way to enable real-time, cost-effective and efficient responses to a variety of human security-related issues.

However, it should be noted that AI is not a panacea. As an inter- and multidisciplinary approach to 'understanding, modelling, and replicating intelligence and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices,' it is effective at carrying out certain tasks but not all.[49] Much depends on the task at hand. For example, AI is very good at finding novel patterns in mass amounts of data.[50] Where humans are simply overwhelmed by the volume of information, the processing power of the computer is able to identify, locate and pick out various patterns. Moreover, AI is also extremely good at rapidly classifying data. Since the 1990s, AI has been used to diagnose various types of diseases, such as cancer, multiple sclerosis, pancreatic disease and diabetes.[51] However, AI is not yet able to reason as humans do, and the technology is far from being a substitute for general human intelligence with common sense.

In short, AI looks to find various ways of using information communication technologies, and sometimes robotics, to aid humans and complete tasks. How the AI is created (its particular architecture) and its purpose (its application) can vary significantly. For the purposes of this chapter, however, the tasks that AI is particularly well suited to, in the human security domain, are related to planning and pattern recognition, especially given big data problem sets. In view of the current considerable capabilities in these areas, it is reasonable to estimate that in the coming years AI will be able to overcome the three constraints on human security-related activities mentioned earlier.

---

[48] One can think of the human security project as a secularized version of the Christian or Jewish notions of 'heaven' or the Islamic idea of 'Jannah', as well as other non-monotheistic religions. The important thing to note is that there is a notion that when one transcends to this place, one is free from all evils, fears, vulnerabilities and needs. Paradise in whatever form is the promise that all ills from the human condition are removed. Human security, likewise, argues for the removal of these same things.
[49] Frankish, K., and Ramsey, W. M. (2014), 'Introduction', in Frankish, K. and Ramsey, W. M. (eds) (2014), *The Cambridge Handbook of Artificial Intelligence*, Cambridge: Cambridge University Press, p. 1.
[50] Sagiroglu, S. and Sinanc, D. (2013), 'Big data: A review', Collaboration Technologies and Systems (CTS), 2013 International Conference, San Diego, CA, https://ieeexplore.ieee.org/document/6567202/, pp. 42–47.
[51] Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., Havel, J. (2013), 'Artificial Neural Networks in Medical Diagnosis', *Journal of Applied Biomedicine*, 11(2), pp. 47–58.

Knowledge

The ability to generate knowledge is no easy feat. Knowledge is subtly different from mere data, which are just an amalgamation of discrete and observable facts or inputs that lack meaning without analysis and context. Only when sets of data are given meaning do they become information, which feeds into and builds knowledge.

There are two obstacles to developing knowledge to tackle human security challenges. The first comes from the considerable amount of data that future generations will generate. Everything from individual output from wearable devices to content created on new communication or social media platforms will saturate the world in an ocean of bits and bytes. There will be a requirement for a way to make these data, from the billions of new devices and millions of new users, intelligible.

Because human security crises can emerge from anywhere and result in various physical or economic social impacts, there will be an urgent need to disentangle discrete flows of data specific to the various vulnerabilities

Second, because human security crises can emerge from anywhere and result in various physical or economic social impacts, there will be an urgent need to disentangle discrete flows of data specific to the various vulnerabilities. Such data flows can be specific to one type of phenomenon, such as extreme weather events prediction,[52] or could even be more diffuse, such as searching and correlating various events or combining datasets to look for indicators of conflict onset or escalation.

AI applications related to search, classification and novel pattern recognition can help to correlate and extract content and meaning from multiple sources. For example, the Early Model-Based Event Recognition using Surrogates (EMBERS) application forecasts key events up to eight days before they happen with a 94 per cent accuracy rate. EMBERS is a '24x7 forecasting system for significant societal events using open source data including tweets, Facebook pages, news articles, blog posts, Google search volume, Wikipedia, meteorological data, economic and financial indicators, coded event data, online restaurant reservations (e.g. OpenTable), and satellite imagery' to forecast events and notify users in real time.[53] This far outstrips current abilities in traditional political science for prediction and explanation of war, where scholars trudge through and manually code content analysis.[54] In the future, it is more likely that scholars or practitioners will use intelligent artificial agents to process real natural language to comprehend text, rather than merely looking for word frequencies and correlations, thereby deepening the capabilities of programs like EMBERS even further.[55]

---

[52] Bauer, P., Thorpe, A., and Brunet, G. (2015), 'The Quiet Revolution of Numerical Weather Prediction', *Nature*, 525, pp. 47–55.
[53] Ramakrishnan, N., (2016), 'EMBERS', http://dac.cs.vt.edu/research-project/embers/.
[54] For example, some scholars do this simply by frequency analysis (or counting the times a particular word or phrase comes up). In one study, Thomas Chadefaux searched Google's database of English-speaking newspapers for a period of 99 years looking for the frequency of various key words associated with 'tensions'. He then correlated this to the Correlates of War dataset to examine if there were any relationship between reported tension escalation and war onset. Unsurprisingly, he found a correlation. Chadefaux, T. (2013), 'Early Warning Signals for War in the News', *Journal of Peace Research*, 51(1), pp. 5–18. More technologically savvy are those scholars such as Mike Thelwall, who utilize a query program to search social media pages, like Twitter, to record the frequency of various topics, the location, the language and even the gender or sentiment of the searched word, phrase or query. See, Thelwall, M. (undated), 'Sentiment Analysis and Time Series with Twitter', http://mozdeh.wlv.ac.uk/resources/TwitterTimeSeriesAndSentimentAnalysis.pdf.
[55] One of the present state-of-the-art AI companies, Google DeepMind, is already working to create a deep neural network that can read real documents, answer complex questions and do so 'with minimal prior knowledge of language structure'. DeepMind's AI is able to perform reasonably well at learning real natural language, as it can presently answer about 64 per cent of queries correctly. Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M. and Blunsom, P. (2015), 'Teaching Machines to Read and Comprehend', http://arxiv.org/pdf/1506.03340v1.pdf.

Another area for the use of AI in human security is health. There are various applications and abilities in this domain but a few are of particular note. First, AI's ability to classify and identify images allows it to recognize patterns more quickly and accurately than people. This has been particularly true in the diagnosis of certain types of cancer.[56]

However, one need not be in a state-of-the-art facility or hospital to receive this type of care. Mobile phones are increasingly being used for bioanalytical science, including digital microscopy, cytometry, immunoassay tests, colorimetric detection and healthcare monitoring. The mobile phone 'can be considered as one of the most prospective devices for the development of next-generation point-of-care (POC) diagnostics platforms, enabling mobile healthcare delivery and personalized medicine'.[57] With advancements in mobile diagnostics, millions more people may be able to monitor and diagnose health-related problems, especially given the estimated increased use in mobile data and devices.

The ability to know what is happening, when and where is the first step in addressing vulnerability

Moreover, with increased connectivity through social media, AI can leverage big data in ways that encourage the uptake of preventive measures. For instance, one application uses machine learning to estimate real-time problematic areas or establishments that may cause food-borne illnesses.[58] This particular application alerts health inspectors in real time to potential outbreaks of food-borne illness so that they may take immediate action. In essence, the ability to know what is happening, when and where is the first step in addressing vulnerability.

Planning

In addition to acquiring and contextualizing knowledge, it is also essential to have the ability to plan an appropriate response. Algorithms related to planning can quickly, reasonably and reliably enable users to carry out complex and multi-stage actions. The need for this facility can be illustrated by examining the UN's average estimated response times for new peacekeeping missions. The UN estimates that when a new crisis emerges – that is, a crisis that involves violence and mass threats to human rights – the estimated response time to plan and field a credible peacekeeping mission is six to 12 months.[59] There are two reasons for this; first, is the strict structure and process of formulating peacekeeping missions.[60] Second, attempts 'to develop better arrangements for rapid deployment have been repeatedly frustrated by austerity and a zero-growth budget'.[61] In short, politicking within the bureaucracy and money constraints limit the UN's ability to act swiftly.

[56] Perry, P. (2016), 'How Artificial Intelligence will Revolutionize Healthcare', Big Think, http://bigthink.com/philip-perry/how-artificial-intelligence-will-revolutionize-healthcare.
[57] Vashist, S. K., Mudanyali, O., Schneider, E. M., Zengerle, R. and Ozcan, A. (2014), 'Cellphone-based devices for bioanalytical sciences', *Journal of Analytical and Bioanalytical Chemistry*, 406, p.3263, doi:10.1007/s00216-013-7473-1.
[58] Sadilek, A., Kautz, H., DiPrete, L., Labus, B., Portman, E., Teitel, J., and Silenzio, V. (2016), 'Deploying nEmesis: Preventing Foodborne Illness by Data Mining Social Media', Association for the Advancement of Artificial Intelligence, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.723.8225&rep=rep1&type=pdf.
[59] Langille, H.P. (2014), 'Improving United Nations Capacity for Rapid Deployment', New York: International Peace Institute, p. 1.
[60] The process for forming a new operation involves 6–7 steps, depending if one considers evaluation as part of the process. The steps involve: initial consultations with a myriad of actors in and outside the UN, technical field assessments, a Security Council resolution, the appointment of senior officials, then planning, deployment and reporting to the Security Council. See, http://www.un.org/en/peacekeeping/operations/newoperation.shtml. The mere fact that planning takes place after all of the political hurdles have been passed increases the length of time to get a mission off the ground. Given that many humanitarian emergencies are ongoing conflicts, much of the information required is already available and could be used to formulate initial operational plans or contingency plans.
[61] Langille (2014), 'Improving United Nations Capacity for Rapid Deployment', p. 1.

Additionally, there are serious problems related to logistics once a mission is approved. The ability to rapidly and reliably estimate, plan and deliver equipment, supplies and services is 'a constant demand across all field operations'. As such, the UN created a Department of Field Support in 2007, whose focus is on developing a standardized approach to forecasting and planning for new operations; human resource planning; supply chain logistics; and evaluative service centres for mission re-tasking and re-planning.[62]

While politics may get in the way of ensuring human security, there are technological solutions that may help. Specifically, advancements in planning algorithms are promising, particularly in emergency response situations. Emergency logistics scheduling is an application that deals with 'the need to identify, inventory, dispatch, mobilize, transport, recover, demobilize, and accurately track human and material resources in the event of a disaster.'[63] Depending upon the type of disaster or crisis, various linear or nonlinear, single- or multi-objective algorithms are presently available for this purpose. These algorithms can identify ideal station points,[64] routing paths for distribution and evacuation,[65] the amount of relief required,[66] and scheduling.[67]

In the coming years, these algorithms are likely to improve, and as they are able to undergo further modification, such as through genetic evolution, learning and adaptation, their abilities will sharpen. Indeed, there is no ostensible reason why the primary objectives of the UN's Department of Field Support could not be met by using planning AI, thereby automating the majority of its tasks. By doing this, time spent on forecasting and planning would fall, and its implementation would cut costs related to human resources and training and remove many redundancies and barriers in supply chain logistics. This would improve the efficiency of any service centres and potentially result in rapid deployment of forces at a reduced cost.

Governments, NGOs and civil society groups can also avail themselves of this AI. NGOs and civil society groups may in fact be best placed to trial these technologies, as they are often not hampered by political obstacles or byzantine bureaucratic rules. They may have greater flexibility to try out new approaches and improve confidence in those ideas. This would help immensely in various human security-related situations, such as complex humanitarian crises – those that are combinations of political and natural disasters – as they could examine vast amounts of data relating to available resources, use satellite imagery or images from surveillance aircraft to map affected terrain, as well as find survivors, and thereby estimate the necessary resource requirements given limitations on time, goods and manpower.

[62] Ibid, p. 13.
[63] Chang, F.-S., Wu, J.-S., Lee, C.-N., and Shen, H.-C. (2014), 'Greedy-search-based multi-objective genetic algorithm for emergency logistics scheduling', *Expert Systems with Applications*, 41(6), p. 2947.
[64] Chang M.-S., Tseng, Y.-L., and Chen, J.-W. (2007), 'A Scenario planning approach for the flood emergency logistics preparation problem under uncertainty', Transportation Research Part E: Logistics and Transportation Review, 43, pp. 737–754.
[65] Zhang, X., Zhang Z., Zhang Y., Wei, D. and Deng, Y. (2013), 'Route Selection for Emergency Logistics Management: A Bio Inspired Algorithm', Safety Science, 54, pp. 87–91.
[66] Zhou, Q., Huang, W., and Zhang, Y. (2011), 'Identifying Critical Success Factors in Emergency Management Using a Fuzzy DEMATEL Method', Safety Science, 49, pp. 243–252.
[67] Sheu, J. B. (2007), 'An Emergency Logistics Distribution Approach for Quick Response to Urgent Relief Demand in Disasters', Transportation Research Part E: Logistics and Transportation Review, 43, pp. 687–709.

Empowerment

As human security has such a broad definition, there are almost limitless ways in which AI can help individuals to be more secure. The key is that such applications empower actors and enable them to make better decisions. Determining how AI can do this without exacerbating existing inequalities or unintentionally creating situations of insecurity is also a consideration. As UN General Assembly Resolution 290 states, 'Human security calls for people-centred, comprehensive, context-specific and prevention-oriented responses that strengthen the protection and *empowerment* of all people and all communities,' acknowledging that it also 'equally considers civil, political, economic, social and cultural rights'.[68]

Empowerment is thus not easily achieved. If all human rights are of equal value, then trade-offs between them are not easily resolved. Furthermore, it is unclear how AI might contribute or detract from such rights

Empowerment is thus not easily achieved. If all human rights are of equal value, then trade-offs between them are not easily resolved. Furthermore, it is unclear how AI might contribute or detract from such rights. For example, AI's ability to find patterns in big data is an asset in diagnosing diseases such as cancer, but it may not be desirable when the pattern that it finds is controversial in some way, such as if it is obviously racist, sexist or extremist. Such patterns may well exist because of the available data or because of existing inequalities or systemic biases in a society. AI could merely be making visible the tyranny of the majority in this situation by classifying particular people, groups or behaviours in various categories.[69] Take, for instance, the Microsoft Twitter bot that within 24 hours of being deployed on Twitter was turned into a racist, sexist and genocidal chatbot due to the amount of these types of phrases being 'fed' to the bot by other Twitter users. Microsoft had to deactivate the bot immediately. When it was accidentally activated a few weeks later, it once again began making inappropriate tweets.[70]

In the most concerning of cases, AI could actually disempower people. This was demonstrated by an algorithm used to predict recidivism rates, which incorrectly scored black defendants in the US along all metrics, such as the likelihood of reoffending or of committing violent acts.[71] These estimates were considered as evidence in sentencing recommendations, and because of systemic race and gender biases against classes of individuals those being sentenced were unfairly and systematically sanctioned.

Thus, it is necessary to interrogate the purpose and effects of AI applications as they relate to empowerment and human security. To facilitate this, one might think of utilizing particular principles as normative guides. An example might be applying something like a Rawlsian principle of justice – which aims to give the greatest benefit to the least advantaged members of society – to AI applications.[72] This would provide a general and high-level principle with which to test various context-specific cases

[68] A/Res/66/290. Italics added.
[69] Take for example the recent case where an image classification algorithm on Google classified images of African-American individuals as gorillas. Google apologized for the incident. BBC News (2015), 'Google apologises for Photos app's racist blunder', 1 July 2015, http://www.bbc.com/news/technology-33347866.
[70] Kosoff, M. (2016), 'Microsoft's Racist Millennial Twitter Bot Strikes Again', Vanity Fair, http://www.vanityfair.com/news/2016/03/microsofts-racist-millennial-twitter-bot-went-haywire-again.
[71] White defendants were estimated to be less likely to reoffend and be of lower risk of committing future crimes (of both a nonviolent and violent nature). Larson, J., Mattu, S., Kirchner, L., and Angwin, J. (2016), 'How we Analyzed the COMPAS Recidivism Algorithm', Pro Publica, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.
[72] Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: Harvard University Press, p. 75. This is otherwise known as the 'difference principle'. The principle has come under much scrutiny for many reasons in its history. For more, see Amartya Sen's essay 'Equality of What?', which lays out many of the problems and a potential solution, Sen, A. (1979), 'Equality of What?', The Tanner Lecture on Human Values, http://tannerlectures.utah.edu/_documents/a-to-z/s/sen80.pdf. Sen's contention is that a difference principle is not sufficient because it does not sufficiently address individual people's needs and capabilities. He instead would see a capabilities approach. For more on the capabilities approach see Nussbaum, M. C. (2011), *Creating Capabilities: The Human Development Approach*, Cambridge, MA: Harvard University Press.

to estimate the likely effects of a given AI application. To succeed, it would require AI application developers to adopt an attitude that reflects both their technical know-how and a consideration of broader social elements. In particularly sensitive applications, such as in those related to potential human rights transgressions, further scrutiny would be warranted to ensure that the data provided were robust and diverse, as well as designed to be mindful of the value of these rights. Recent work by the US Federal Trade Commission and the White House on the need for further regulation of big data and algorithmic-based decisions, such as through best practices, codes of conduct and even existing or new laws, is also important.[73]

Equitable, transparent and accountable

From a policy standpoint, it is essential to know what data are used, an AI model's guiding assumptions, as well as the kinds of practices that developers utilize

Ultimately, as more data are used to influence decisions, and as algorithms are increasingly utilized to shape, guide or make these decisions, humans must be vigilant in asking for transparency, accountability, equity and universality from these applications. These are all elements of ensuring a just distribution and accounting of the benefits of AI. From a policy standpoint, it is essential to know what data are used, an AI model's guiding assumptions, as well as the kinds of practices that developers utilize. An understanding of these components will allow for accurate estimates of the likely effects of an AI application.

As it is known that technology is not value neutral, but is in fact a human creation guided by (often implicit) particular values, policymakers ought to ensure that such technologies and their benefits are accessible to everyone in an open source format. Human security is not for an elite few, and so the capabilities of AI must be within everyone's grasp.[74] When it comes to applications related to disaster relief, conflict prevention, human rights protection and justice, it is imperative that wider schemes of data sharing are employed by individuals, groups, NGOs and governments. However, it is simultaneously imperative that data, through sharing and acquisition, are also protected to the greatest possible extent. Health, in particular, is one immediate area of focus for privacy, transparency and accountability policies, best practices and regulation.[75]

These concerns are important to human security activities. Take, for instance, the movement to require biometric identification to receive humanitarian aid. While the intention is to track individuals to reduce fraud, these data could also be used for political oppression. UNHCR, for example, argues for increased biometric identification, but these data are shared by a variety of actors for multiple purposes.[76] Yet there

[73] United States of America, Executive Office of the President (2016), 'Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights', https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf; The Federal Trade Commission (2016), 'Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues', https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report.
[74] The OpenAI Initiative is one attempt at making AI open source and available to all. Unfortunately, however, the amounts of data required for robust AI is usually beyond the reach of private individuals and remains in the hands of governments and corporations. https://openai.com/about/.
[75] Google DeepMind, for example, has instituted a group of independent reviewers to oversee DeepMind Health applications. The reviewers are trusted public figures, and form a multi-stakeholder perspective. They are not paid, have no conflicts of interest, and are put in place to scrutinize DeepMind Health's 'work, its handling of data, investigate data sharing agreements, understand the product roadmaps and critique the handling of health data' from the UK's National Health Service, https://deepmind.com/health/independent-reviewers.
[76] Ismail, Y. (2006), 'Fingerprints Mark New Direction in Refugee Registration', UNHCR, http://www.unhcr.org/456ede422.html.

is no discussion of data protection, and there is a gap with regard to policy guidance or compliance. Thus, in situations of complex humanitarian disasters, where refugee safety is certainly of concern, regulations need to be established.[77]

In short, AI that enables human security must, by its very nature, ensure that it is aimed at minimizing human insecurity, maximizing human empowerment, and is as equitable, transparent and accountable as possible. The consequences of algorithms misclassifying or failing to plan appropriately could be catastrophic. Therefore, good policy, regulation and accountability measures need to be in place. These may range from pre-emptively instituting a set of best practices to remedial or coercive measures after the fact. Whatever the situation, humanity must not walk into a future increasingly influenced by AI and claim the equivalent of an AI 'Twinkie Defense'.[78] Therefore, AI that is sensitive to context, vulnerability and capacity-building, but guided by good judgment, foresight and principles of justice would be most beneficial for all.

[77] For instance, encountering a biometric fingerprint scanner in Malaysia, one refugee from Myanmar was quoted as saying: 'I don't know what it is for, but I do what UNHCR wants me to do.' This sort of trust in the UN system's ability to protect refugees and their digital information is questionable. Cited in Currion, P. (2015), 'Eyes Wide Shut: The Challenges of Humanitarian Biometrics', IRIN, http://www.irinnews.org/opinion/2015/08/26/eyes-wide-shut-challenge-humanitarian-biometrics. The same article cites the case of some 84,000 biometric UK prisoner records having been compromised when they were left on an unencrypted USB stick in an unlocked drawer.

[78] The so-called 'Twinkie Defense' was used by Dan White's defence lawyer to argue that when he murdered Harvey Milk he was suffering from mental impairment due to too much sugar consumed from eating Twinkies and was thus not responsible for his act, https://www.law.cornell.edu/wex/twinkie_defense.

4. The Economic Implications of Artificial Intelligence

Kenneth Cukier

Modern AI

The idea of using machines to recreate human work has a long lineage. In the West, it dates back to the ancient Greek god Hephaestus, said to have created mechanical servants and a man of bronze, Talos, to protect Crete from invaders. Aristotle noted: 'When the loom spins itself and the lyre plays by itself, man's slavery will be at an end.'[79]

Yet the idea that a computer might achieve human-like 'intelligence' received its most important push by the mathematician and computer scientist Alan Turing in 1950.[80] In his essay 'Computing machinery and intelligence', published in the journal *Mind,* he outlined the basis of a 'learning machine' that, like a child, was given some basic logic and then taught over time. 'We may hope that machines will eventually compete with men in all purely intellectual fields,' he wrote.

In subsequent decades, AI made progress but never matched the ambitions of its early pioneers. Over the past decade, however, something extraordinary happened: techniques that previously hadn't worked very well started to show impressive advances because processors became more powerful and there were vastly more data to train the algorithms (hence the idea of 'big data').[81]

Today, machine learning, and in particular a sophisticated variant called 'deep learning', are at the heart of many applications that may seem almost magical. Video-recognition technology can lip-read speech; character-recognition systems can decipher written script; photo software identifies people in pictures; voice-recognition systems respond to speech; and real-time computer translation is moving from the research laboratory to commercial products. And this is to say nothing of self-driving cars and applications for medical diagnosis. AI is behind the rise of the robots too. The huge advances in robotics in recent years are largely attributable to machine-learning algorithms.

It is successes like these that call to mind Turing's hope that 'machines will eventually compete with men', or echo Aristotle's vision that autonomous things may end 'man's slavery'. Yet advances in AI are also compelling economists and others to consider its impact in three main areas: productivity and output; employment; and international trade and development.

The changes arising from the development and application of AI may be on par with those resulting from electrification or of computing

Impact on productivity and output

There is a good chance that AI will bring about a boom in productivity. The changes arising from the development and application of AI may be on par with those resulting from electrification or of computing. One of the pioneers of the internet, Scott Bradner of Harvard University, once described the internet revolution as 'the platform for all subsequent revolutions'.[82] The same can be said of AI. It is a general-purpose technology, and as seen with computers, the innovations that take place in AI will spark revolutions in every domain that it touches.

[79] Aristotle as quoted in Partridge, D. and Hussain, K. M. (1992), *Artificial Intelligence and Business Management*, New Jersey: Ablex Publishing Corporation.
[80] Turing, A. M. (1950), 'Computing Machinery and Intelligence', *Mind* 59(236), pp. 433–460, http://phil415.pbworks.com/f/TuringComputing.pdf.
[81] Mayer-Schönberger, V. and Cukier, K. (2013), *Big Data: A Revolution That Transforms How We Live, Work and Think*, London: John Murray.
[82] Author's interview with Scott Bradner, 1995.

Several investment banks and consultancies have attempted to forecast the economic value of applying AI to existing activities by industry and geography. Bank of America (BoA) Merrill Lynch in 2015 said it expected that in 10 years, robotics and AI would produce annually an 'impact value' of between $14 trillion and $33 trillion,[83] potentially including, at the upper end of this range, $8–$9 trillion in cost reductions across manufacturing and healthcare, $9 trillion in employment cost savings, and some $1.9 trillion in efficiency gains by autonomous cars and drones.

By a different methodology, in 2013 the McKinsey Global Institute came to the range of $10 trillion to $25 trillion per year by 2025 for robotics, AI and data-intensive activities like the 'internet of things', industrial sensors, etc.[84] McKinsey further estimates, according to a study published in 2017, that automation could raise productivity growth worldwide by 0.8–1.4 per cent per year.[85] A 2016 report by the Analysis Group, funded by Facebook, estimated the 'reasonable range' of economic impact of AI over the next 10 years at between $1.49 trillion and $2.95 trillion.[86] (Note that the period covered by the Analysis Group is 2016–26, whereas the forecasts by BoA and McKinsey only begin in 2025.)

A report by Accenture in 2016, based on economic modelling by Frontier Economics, estimated that the widespread use of AI in society and business had the potential to double the annual economic growth rate by 2035 in the dozen developed economies studied (the US, 10 Western European countries and Japan). The report forecasts, for example, that in the case of the US, absorption of AI in the economy would increase the rate of growth in gross value added (GVA) – a close approximation to gross domestic product (GDP) – from a baseline 2.6 per cent to 4.6 per cent in 2035, equivalent to an additional $8.3 trillion GVA per year.[87]

In 2017, moreover, PwC forecast that GDP worldwide could be as much as 14 per cent higher in 2030 because of AI technologies, which it valued as potentially contributing some $15.7 trillion to the global economy. The majority of the gains would, in its assessment, come from retail, financial services and healthcare in terms of greater productivity, enhanced products and higher demand.[88]

The competition among these institutions to produce eye-watering estimates as to AI's potential value appears to match the rivalry in the AI industry itself. Their figures vary widely, but what is clear is that the sums involved are vast, and that a huge amount of value is predicted to be unleashed.

How might this happen in practice? Consider just one area of AI: self-driving cars. If all taxis, ride-shares and long-haul delivery vehicles are autonomous, then a huge amount of labour costs will be saved (though some of that saving will be spent on the

[83] Ma, B., Nahal, S. and Tran, F. (2015), 'Robot Revolution – Global Robot & AI Primer', Bank of America Merrill Lynch, 16 December 2015.
[84] Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P. and Marrs, A. (2013), *Disruptive technologies: Advances that will transform life, business, and the global economy*, McKinsey Global Institute, http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/disruptive-technologies.
[85] Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P. and Dewhurst, M. (2017), *A Future That Works: Automation, Employment, and Productivity*, McKinsey Global Institute, https://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works.
[86] Chen, N., Christensen, L., Gallagher, K., Mate, R. and Rafert, G. (2016), *Global Economic Impacts Associated with Artificial Intelligence*, Analysis Group, http://www.analysisgroup.com/uploadedfiles/content/insights/publishing/ag_full_report_economic_impact_of_ai.pdf.
[87] Purdy, M. and Daugherty, P. (2016), *Why Artificial Intelligence is the Future of Growth*, Accenture, https://www.accenture.com/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.pdf.
[88] PwC (2017), *Sizing the prize: What's the real value of AI for your business and how can you capitalise?*, https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf.

AI technology that makes the transformation possible). The vehicles will, moreover, use less fuel, emit fewer pollutants and have fewer accidents, and human former drivers will be freed to do other economically and socially useful things. Accordingly, McKinsey estimates that the potential economic impact of autonomous cars and trucks alone could be in the wide range of $200 billion to $1.9 trillion per year by 2025.[89] Society will gain the value of transport services at less cost – and perhaps therefore increase its use of transport because it is now cheaper. Meanwhile, the savings can be allocated to more productive uses, such as investment in new ideas.

Next, consider a specific job: a human welder in a car factory. Today in the US, a worker in this job earns around $25 per hour, including benefits. The equivalent operating cost per hour for a robot is around $8, including installation, maintenance and depreciation. By 2030 the cost could fall to $2 per hour, estimates the Boston Consulting Group in a study published in 2015.[90] Even that sounds conservative: it means the price would halve twice in 15 years. From the experience of price–performance drops due to computing technology, a fall of an order of magnitude seems more likely, i.e. to 80 cents per hour.[91] In any case, the cost of labour to produce vehicles declines.

The more interesting benefit – and the sine qua non of AI – is in doing things that no group of humans could reasonably hope to achieve. Business is already moving towards that vision

These two examples are just the beginning. They assume that AI and robotics do the same thing as people, only more cheaply. The more interesting benefit – and the *sine qua non* of AI – is in doing things that no group of humans could reasonably hope to achieve. Business is already moving towards that vision.

An example is Otto, a large German online retailer (unrelated to the American self-driving trucking company of the same name). Its data showed that customers were more likely to return items if they arrived three days or more after ordering, meaning that the company needed to shorten the shipping time to improve customer service and protect its margins. To do this, Otto deployed a predictive model using a deep learning algorithm by a company called Blue Yonder that was originally developed for analysing experiments at CERN, the European Organization for Nuclear Research. The algorithm analysed some 200 variables (such as season, weather, colour, style, size, etc.) to predict what shoppers would buy, a full month before the purchases actually happened, enabling Otto to have stocks in place so it could dispatch goods immediately after orders were placed. The system is fully automated. In 2017 Otto's automated procurement system was purchasing 200,000 products a month completely on its own. The project was a triple win: it got most merchandise to customers within two days, it reduced returns by 2 million items a year (thereby saving Otto millions of dollars) and it also helped the environment, since fewer packages were sent back.[92] It would be a significant, if not impossible, endeavour to assemble a team of humans to crunch through 200 discrete variables to buy millions of items in anticipation of customer orders. But an AI algorithm can do this.

It is difficult to calculate the increase in productivity resulting from AI. Productivity tracks output per worker. Total factor productivity (TFP) attempts to measure the improvement from technology by considering output in terms of a weighted average

[89] Manyika et al. (2013), *Disruptive technologies*.
[90] Sirkin, H., Zinser, M. and Rose, J. (2015), 'How Robots Will Redefine Competitiveness', The Boston Consulting Group, 23 September 2015, https://www.bcgperspectives.com/content/articles/lean-manufacturing-innovation-robots-redefine-competitiveness/.
[91] See for example U.S. Bureau of Labor Statistics (2015), 'Long-term price trends for computers, TVs, and related items', https://www.bls.gov/opub/ted/2015/long-term-price-trends-for-computers-tvs-and-related-items.htm.
[92] *The Economist* (2017), 'Automatic for the people: How Germany's Otto uses artificial intelligence', 12 April 2017, https://www.economist.com/news/business/21720675-firm-using-algorithm-designed-cern-laboratory-how-germanys-otto-uses.

of capital and labour. The 'residual' is innovation – or, as economists quip, 'a measure of our ignorance'. These metrics fail to capture the value that AI creates for two reasons. First, the output component is based on GDP, and GDP only measures monetary transactions. If the service is free, it is invisible as output.[93] Second, and worse from a measurement perspective, if the free, AI-infused service replaces a paid-for service, GDP actually goes down – and with it *measured* productivity.

This is a very real problem for AI. To take one example, Google upgraded its web-translation service in the autumn of 2016 from using classic statistical techniques to relying on deep learning.[94] This made the translations considerably more accurate. If a user chose to rely on Google's free translations rather than pay someone to translate a document, the economic value of AI would be clear – but it would have a negative effect on GDP, and thus on productivity. The problem is compounded with each upgrade, because the performance improvement is invisible to economic metrics.

The biggest advantage of AI is in performing tasks that are entirely new, or at least far beyond reasonable human capacity. Yet these are the very tasks prone to mismeasurement

Moreover, the biggest advantage of AI is in performing tasks that are entirely new, or at least far beyond reasonable human capacity. Yet these are the very tasks prone to mismeasurement. To paraphrase the economist Robert Solow, the value of AI is 'everywhere but in the productivity statistics'.[95]

In the case of AI, some economists are questioning whether this time really *is* different. Can we expect a breakaway in productivity arising from AI? Or will it produce only a modest blip? In a paper issued by the Brookings Institution in 2016, David Byrne of the US Federal Reserve Board and colleagues noted that, notwithstanding growing mismeasurement, productivity has lagged over the period 1995–2004 precisely when IT deployment soared.[96] The implication is to not expect too much from the latest round of digital innovation.

There is a view, put forward notably by Robert Gordon of Northwestern University, that AI and robotics will bring about only minor advances in TFP compared with the extraordinary gains of the early 20th century.[97] This view, though popular, needs to be tempered. Just because earlier productivity gains from technologies such as railways and the telephone are more evident to economists does not refute the assertion that something unique is happening with AI. The assessment may also be the victim of bad timing. Professor Gordon began promoting his research findings in 2012, the same year that the foundational papers in deep learning were published,[98] suggesting that his conclusions may not have taken account of the latest breakthroughs in AI technology.

---

[93] On the failure of GDP statistics to capture the value of free online services, see: Brynjolfsson, E. and Oh, J. H. (2012), 'The Attention Economy: Measuring the Value of Free Goods on the Internet', http://ide.mit.edu/research-projects/attention-economy-measuring-value-free-goods-internet.
[94] Lewis-Kraus, G. (2016), 'The Great AI Awakening', *New York Times Magazine*, 14 December 2016, https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.
[95] By Solow's eponymous paradox, coined in 1987, 'You can see the computer age everywhere but in the productivity statistics.'
[96] Byrne, D., Fernald, J. G. and Reinsdorf, M. B. (2016), *Does the United States have a Productivity Slowdown or a Measurement Problem?*, Washington, DC: Brookings, https://www.brookings.edu/bpea-articles/does-the-united-states-have-a-productivity-slowdown-or-a-measurement-problem/.
[97] Gordon, R. J. (2016), *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*, Princeton, NJ: Princeton University Press.
[98] Most notably, Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012), 'ImageNet Classification with Deep Convolutional Neural Networks', *Advances in Neural Information Processing Systems* 25 (NIPS 2012); and Le, Q. V., Ranzato, M., Monga, R., Devin, M., Corrado, G. S., Chen, K., Dean, J. and Ng, A. Y (2012), 'Building high-level features using large scale unsupervised learning', 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (2012), pp. 8595–98.

Meanwhile, from a highly mathematical econometrics perspective, William Nordhaus of Yale University does not believe that AI will produce an 'economic singularity' – a point at which technology produces exponential growth as far as the eye can see.[99]

One way to reconcile the apparently poor productivity figures of recent years is to look at firm-level performance. There has been a marked phenomenon of 'the best versus the rest' – i.e. a tremendous boom in productivity by the very best firms (those in the 90th and 99th percentile of performance) and productivity stagnation by the rest. The outsized performance gains – attributed in part to technology and automation – are camouflaged by the broader, middling economy. For instance, according to the Organisation for Economic Co-operation and Development (OECD), between 2001 and 2007 labour productivity at the 'global frontier' (referring to the top 5 per cent of firms in terms of multi-factor productivity) grew by 4–5 per cent per year, compared with annual growth averaging just 1 per cent by the rest.[100] The firms at the forefront of their industries are more likely to be pioneering users of AI.

Andrew Haldane, the chief economist at the Bank of England, has identified the same trend in an analysis of 300,000 British firms.[101] The gross value-added per person was seven times greater among firms in the 99th percentile than among those at the median. The 'winner-take-all' profile is brutal: those at the 99th percentile are three times more productive than those at the 90th percentile – already among the very best-run companies.

In light of this, some economists embrace the idea that rather than fear a world where we have AI, we should be alarmed by a world where we have too little of it.

What is clear is that human welfare and living standards are likely to improve as people can enjoy more goods and services – although the gains may be very unequally distributed

To measure the performance of AI, a panel convened by the US National Academy of Sciences released a report in April 2017 that recommended, among other things, the establishment of an 'AI progress index'.[102] Comparable to the consumer price index (CPI) as a means of tracking inflation, the index would track progress on specific AI and machine learning technologies. This is a fascinating idea, even if it is prone to statistical wrinkles, like CPI itself.

Ultimately, however, no one can say for sure what the value of AI will be for the economy. What is clear is that human welfare and living standards are likely to improve as people can enjoy more goods and services – although the gains may be very unequally distributed. By improving efficiency and lowering costs, AI will likely mean more output and thus perhaps lower prices for consumers. This may free up capital to be deployed more efficiently elsewhere. However, to return to the examples above, what will this mean for the car driver or welder?

[99] Nordhaus, W. D. (2015), 'Are We Approaching An Economic Singularity? Information Technology and The Future of Economic Growth', *National Bureau of Economic Research Working Paper*, No. 21547, http://www.nber.org/papers/w21547.pdf.
[100] Andrews, D., Criscuolo, C. and Gal, P. N. (2016), 'The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy', *OECD Productivity Working Paper*, No. 5, www.oecd-ilibrary.org/economics/the-best-versus-the-rest_63629cc9-en.
[101] Haldane, A. G. (2017), 'Productivity puzzles', speech given at London School of Economics, London, 20 March 2017, www.bankofengland.co.uk/publications/Documents/speeches/2017/speech968.pdf.
[102] National Academies of Sciences, Engineering, and Medicine (2017), *Information Technology and the U.S. Workforce: Where Are We and Where Do We Go from Here?* Washington, D.C.: The National Academies Press, https://www.nap.edu/catalog/24649/information-technology-and-the-us-workforce-where-are-we-and.

Impact on jobs and wages

The orthodox view is that AI will lead to a 'jobs apocalypse' whereby mass unemployment becomes the norm. Numerous reports and books have sounded the alarm, predicting massive job losses.[103] The historian Yuval Noah Harari goes even further in his book *Homo Deus*, arguing that AI will render a vast swathe of humanity 'useless'.[104]

The potential solutions put forward range from taxing robots to decoupling incomes from jobs. There have also been calls for more substantial interventions, such as the state providing a 'universal basic income' (UBI) to redistribute the gains from the owners of the means of production who deploy AI to the vast army of the unemployed who are AI's victims. (The Marxist language here is deliberate, in so far as there are parallels to draw between the nascent AI economy and the industrial revolution of Karl Marx's day.)

Economic history shows that automation creates new jobs around the new processes, and that these new jobs still require people

However, some economists see the view that there is a fixed amount of work to go around, and that if robots take it, humans will be left with nil, as a 'lump of labour fallacy'. Economic history shows that automation creates new jobs around the new processes, and that these new jobs still require people. 'Technology eliminates jobs, not work,' noted a US government report on automation published back in 1966.[105]

It is a fractious and essential debate. On one side are tech-minded people close to the advances, on the other are economists who are more removed but who bring history and a longer view. This chapter leans towards the optimistic outlook, but does not make a case for it. Instead, it will first present the argument that AI may obliterate jobs, and then put forward reasons why it may not. It will close on an area of consensus – that wages may suffer.

The idea of a jobs apocalypse was sparked by a study in 2013 by Carl Benedikt Frey and Michael Osborne of Oxford University. They concluded that as many as 47 per cent of jobs in the US are susceptible to automation over the next several decades.[106] The corresponding share for the UK is 35 per cent.[107] The paper notably does not say that jobs *will* disappear or give a specific timeframe; rather that they are susceptible over some decades.

The Bank of England, in an informal exercise using the same methodology, put potential job losses even higher, at around half the British workforce.[108] A study by PwC estimated that some 30 per cent of British jobs are vulnerable to automation from AI and robotics by the early 2030s; the comparable estimates for the US and Germany are 38 per cent and 35 per cent, while that for Japan is somewhat lower, at 21 per cent.[109]

---

[103] Books include: Avent, R. (2016), *The Wealth of Humans: Work, Power, and Status in the Twenty-First Century*, New York, NY: St Martin's Press; Ford, M. (2015), *Rise of the Robots: Technology and the Threat of a Jobless Future*, New York, NY: Basic Books; Brynjolfsson, E. and McAfee, A. (2014), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York, NY: W. W. Norton; Kaplan, J. (2015), *Humans Need Not Apply: A Guide to Wealth and Work in the Age of Artificial Intelligence*, New Haven, CT: Yale University Press; Chace, C. (2016), *The Economic Singularity: Artificial Intelligence and the Death of Capitalism*, London: Three Cs.
[104] Harari, Y. N. (2016), *Homo Deus: A Brief History of Tomorrow*, London: Harvill Secker.
[105] From Bowen, H. R. and Magnum, G. L. (1966), 'Report of the National Commission on Technology, Automation, and Economic Progress: Volume I', Washington, DC: United States Government. As quoted in Autor, D. H. (2015), 'Why Are There Still So Many Jobs? The History and Future of Workplace Automation', *Journal of Economic Perspectives*, 29(3), pp. 3–30, http://dx.doi.org/10.1257/jep.29.3.3.
[106] Frey, C. B. and Osborne, M. (2013), 'The Future of Employment: How Susceptible are Jobs to Computerisation?', Oxford: Oxford Martin School, http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf.
[107] Deloitte LLP (2014), *Agiletown: the relentless march of technology and London's response*, London: Deloitte LLP, https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/uk-futures/london-futures-agiletown.pdf.
[108] Haldane, A. G. (2015), 'Labour's Share', speech given at Trades Union Congress, Bank of England, London, 12 November 2015, https://www.bankofengland.co.uk/speech/2015/labours-share.
[109] PwC (2017), 'Up to 30% of existing UK jobs could be impacted by automation by early 2030s, but this should be offset by job gains elsewhere in economy', PwC Blog, 24 March 2017, http://pwc.blogs.com/press_room/2017/03/up-to-30-of-existing-uk-jobs-could-be-impacted-by-automation-by-early-2030s-but-this-should-be-offse.html.

AI is being introduced to many different industries at the same time, unlike during other periods of jobs upheaval such as the early industrial revolution

One fear is that job destruction will happen faster than new job creation. AI is being introduced to many different industries at the same time, unlike during other periods of jobs upheaval such as the early industrial revolution. Then, for example, if a weaver was put out of work through the introduction of 'power looms', there were other industries to turn to. With AI, many sectors may all shed jobs at once. Other than retraining as an AI engineer – an option open only to a few – there will be no place to turn.

The emblematic case is that of taxi drivers and truckers once self-driving vehicles hit the road. The debate continues as to when that may happen; early estimates by some in the industry, at around 2020, are unrealistic. It may take considerably longer for full 'level-five' autonomy to be viable, where the machine is in complete control. Nevertheless, the day will likely come. What will happen to employment, considering that there are 3.5 million truck drivers, 230,000 taxi drivers and 500,000 school bus drivers in the US alone?[110] For this reason, many people believe that AI will create a massive upheaval in the jobs market.

The threat to jobs, though real, might not be so dire. Despite the rapid research advances, it will take time for new applications to be adopted, meaning that the economy will have time to adapt. And alongside AI technologies, 'human' skills of personal interaction, teamwork and emotional intelligence will continue to be needed in the labour force. Moreover, the crux of AI technology is not that it recreates what people can do, but that it may exceed that capacity millions-fold, performing tasks that no group of humans could possibly achieve – so there is a case to be made that rather than destroy jobs, AI may in fact liberate people, for whom the new work may be more fulfilling. Economic history offers mixed lessons in this regard, but mostly falls on the side of optimism.

Forecasts of mass unemployment arising from AI are open to criticism. For example, the Frey and Osborne study noted above overstates the problem by dint of its methodology, including its focus on jobs not tasks. Among the occupations that it concluded were most liable to automation were models, baristas, cooks and manicurists. These are the very jobs that people associate with a human touch – which might actually command a premium in a world teeming with bots and algorithms.

Jobs are comprised of tasks, which themselves vary in the degree to which they can be automated. For example, much of a lawyer's job can be done by software, like finding the right precedent and constructing arguments.[111] But among the valuable activities that lawyers fulfil is to ease the anxiety of clients in life-altering situations such as divorce, criminal charges or inheritance disputes. The work is about empathy, not just answers. Strikingly, in the US, as legal software has been introduced, since 2000, the number of law clerks and paralegals has grown – and at a faster rate than the overall workforce – rather than declined.[112]

A task-based analysis is subtler. According to a 2016 McKinsey study of more than 2,000 work activities across 800 jobs, around half of the activities in the global economy could be automated with current technology. About 60 per cent of occupations have at least

[110] American Trucking Associations (2017), 'Trucking Industry Revenues Were $676.2 Billion in 2016', 14 August 2017, http://www.trucking.org/article/Trucking-Industry-Revenues-Were-$676.2-Billion-in-2016; U.S. Bureau of Labor Statistics (2015), 'Taxi Drivers, Ride-Hailing Drivers, and Chauffeurs', https://www.bls.gov/ooh/transportation-and-material-moving/taxi-drivers-and-chauffeurs.htm; U.S. Bureau of Labor Statistics (2017), 'Occupational Employment and Wages, May 2017: 53-3022 Bus Drivers, School or Special Client', https://www.bls.gov/oes/current/oes533022.htm.
[111] Susskind, R. and Susskind, D. (2016), *The Future of the Professions: How Technology Will Transform the Work of Human Experts*, Oxford: Oxford University Press.
[112] Bessen, J. (2016), 'The Automation Paradox', *The Atlantic*, 19 January 2016, https://www.theatlantic.com/business/archive/2016/01/automation-paradox/424437/.

a third of tasks that could be automated, but fewer than 5 per cent of jobs can be entirely automated, the report noted.[113] Moreover, an OECD tasks-based study, also published in 2016, found that, on average, just 9 per cent of jobs in the organization's 21 (mostly advanced) economies are automatable.[114]

While still troubling, the OECD's 9 per cent and McKinsey's 5 per cent assessments regarding elimination of jobs through automation look a lot less dramatic than Oxford's 47 per cent and PwC's 35 per cent forecasts.

The implication is that jobs won't disappear, but they will be different. Recall T. S. Eliot's *Preludes* from 1920:

> And at the corner of the street
> A lonely cab-horse steams and stamps.
>
> And then the lighting of the lamps.[115]

London no longer has lamp-lighters; those so inclined found other ways to make a living. As for the cab-horsemen, they became taxi drivers (and, more recently, 'partner-drivers' for services such as Uber). Indeed, in a study of the US Census Bureau's occupations between 1950 and 2010, only one job – elevator operator – was removed because of automation. Other jobs continue to be performed, even if the nature of that work has changed because tasks were partially automated.[116]

Counterintuitively, automation can sometimes actually increase employment in the industries it touches. This is because the technology lowers the price of a good or service, so people consume more.

In the 19th century, around 98 per cent of the labour needed to weave a yard of cloth became automated. But the number of workers employed in weaving increased during that period. The reason for this was that the price of cloth fell dramatically because of automation, so people bought much more cloth. Individuals who might only have owned a single set of clothing now could have several. Curtains appeared at windows. Furniture was more commonly upholstered. The result was more jobs alongside more automation.[117]

Other industries had similar experiences, notably the growth in bank representatives in the US as cash machines became more prevalent there.[118] And as already noted, legal assistant jobs have grown even as law firms have introduced software to automate some tasks: it reduces costs and makes employees more efficient. Over the past 15 years, automation has created around four times as many jobs in the UK as it has destroyed, according to a report published by Deloitte in 2015.[119]

---

[113] Chui, M., Manyika, J. and Miremadi, M. (2016), 'Where machines could replace humans—and where they can't (yet)', *McKinsey Quarterly*, July 2016, http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet.
[114] Arntz, M., Gregory, T. and Zierahn, U. (2016), 'The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis', *OECD Social, Employment and Migration Working Papers*, No. 189, Paris: OECD, http://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5jlz9h56dvq7-en.
[115] Eliot, T. S. (1920), 'Preludes' in *Prufrock and Other Observations*, New York, NY: A.A. Knopf, www.bartleby.com/198/3.html.
[116] Bessen, J. (2016) 'How Computer Automation Affects Occupations: Technology, Jobs, and Skills', *Boston University School of Law, Working Paper* No. 15-49, https://www.bu.edu/law/working-papers/how-computer-automation-affects-occupations-technology-jobs-and-skills/.
[117] Bessen (2016), 'The Automation Paradox'.
[118] Ibid.
[119] Deloitte LLP (2015), *From Brawn to Brains: The impact of technology on jobs in the UK*, London: Deloitte LLP, https://www2.deloitte.com/uk/en/pages/growth/articles/from-brawn-to-brains--the-impact-of-technology-on-jobs-in-the-u.html.

Another reason why work will not disappear entirely is that there are many jobs in which human engagement is indispensable. Education, healthcare, sales and other areas in which empathy and social skills are critical. People adapt their skills to where they have comparative advantage over software code and computer chips.

For example, if AI systems can detect the propensity of certain diseases, there will be a large share of the population concerned about taking steps to lower their chances of suffering from the ticking time-bomb inside their bodies. This will mean changes in lifestyle, diet, exercise, stress levels, etc. People may be eager to have 'health coaches' support them, just as they have turned to therapists, nutritionists, executive coaches, birth coaches and personal trainers over the past several decades – jobs that largely did not exist a century ago. Already, for example, the profession of 'death doulas' has emerged.

Although robots could perform those tasks, it is doubtful that they would create the meaningful bond that most people want. Unsurprisingly, employment growth in the face of automation has been happening in jobs with substantial social skills, such as nursing, teachers and care workers, according to a study by Deloitte.[120] One AI entrepreneur, Chris Boos of Arago in Germany, believes the day will soon arrive when nurses are paid more than doctors – i.e. emotional nous will trump analytical skills. This is a deliberate provocation, but it makes the point.[121]

Humans still outperform machines when judgment, not just mechanized work, is called for. Today's AI systems are burdened by needing some explicit instruction, even if newer techniques enable the algorithm to infer answers more subtly than ever. Where clear instructions are needed, the technology comes up against what economists call 'Polanyi's Paradox'. This is the role of tacit knowledge, summed up by the phrase 'We know more than we can tell.' Because many tasks cannot be clearly delineated and communicated, we will still need humans to perform them.[122]

In sum, the fear of substantial unemployment because of AI is legitimate, but it is not clear how suddenly it may happen, or how deep it may go. A key question is timing. Will new jobs be created fast enough to replace the ones that go away? And even if jobs do still exist, the types of jobs, their quality, and the skills necessary to perform them are, critically, not fully understood. This is not reassuring for people who have trouble adapting to new job requirements or acquiring new skills. The mismatch between abilities and jobs may be considerable.

Distinct from the jobs question is the issue of wages. When the returns on capital are greater than the returns on labour, firms invest money in machines to perform tasks rather than hire staff. The only way people remain competitive therefore is to work for less. As Jason Furman, chairman of the US Council of Economic Advisers under President Obama, put it in 2016: 'The traditional argument that we do not need to worry about the robots taking our jobs still leaves us with the worry that the only reason we will still have our jobs is because we are willing to do them for lower wages.'[123]

---

[120] Stewart, I., De, D. and Cole, A. (2015), *Technology and people: The great job-creating machine*, London: Deloitte LLP, http://www2.deloitte.com/uk/en/pages/finance/articles/technology-and-people.html.
[121] Boos, C. (2015), from AI panel at 24 Hours conference, Deutsche Telekom, Hohenkammer, Germany, May 2015.
[122] Autor, D. (2014), 'Polanyi's Paradox and the Shape of Employment Growth', *National Bureau of Economic Research Working Paper*, No. 20485, www.nber.org/papers/w20485.
[123] Furman, J. (2016), 'Is This Time Different? The Opportunities and Challenges of Artificial Intelligence', Remarks at 'AI Now: The Social and Economic Implications of Artificial Intelligence Technologies in the Near Term' at New York University, New York, 7 July 2016, https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160707_cea_ai_furman.pdf.

Over time, wages tend to readjust upwards as the inventions that shake the labour market make their way through the economy, and public policy works to redistribute the gains. But this process can be extremely protracted. British workers, for instance, did not see substantial real wage gains during the early industrial revolution until the 1840s, some 60 years after the labour upheaval began.[124]

A scenario for how this might play out in the AI economy can be seen in one economic trend since 2000: offshoring. In this case, jobs have not been replaced by robots but by lower-cost workers in developing countries. This resulted in lay-offs, and put downward pressure on wages in advanced economies (and arguably planted the seeds for the populist surge in recent years).[125]

There is some initial empirical evidence of the impact of automation on jobs. A study of industrial robots in local US economies between 1990 and 2007, published in 2017, found that for every robot per thousand employees, 6.2 workers lost their jobs, and wages dipped by an average of 0.7 per cent in those communities. That said, the impact was less severe on a national level since other jobs were created elsewhere.[126]

The experiences of offshoring and industrial robots provide a glimpse of things to come, since it shows how lower-cost alternatives to labour affect the job market. If employment goes to a machine, it is hard for a human to compete. If jobs go overseas, workers in one country lose out but those in the other gain. However, even this model of international labour markets is likely to change because of AI.

Impact on international trade and development

AI will infiltrate not just advanced economies. A similar wave is poised to hit developing countries, where the effects may be even harsher. This is because AI throws into question the traditional development model, based on low-cost labour, by which poor countries typically achieve economic growth.

The 20th century 'catch-up' model of development relies on international trade. Countries start by exporting low-value goods to richer countries, which no longer bother to make cheap products at home. As the lower-income country becomes more skilled, it moves up the value chain to produce higher-value goods. So a country may go from basic inputs like chemicals and steel to more skilled, manufactured products like cars and electronics, a process that normally takes decades.

Poor countries meanwhile use policy levers such as inviting in foreign capital, with conditions. For example, foreign firms looking to invest in a developing economy are required to set up joint ventures with local firms, adhere to quotas for hiring locals in skilled tasks, agree to technology transfer, etc. The local country can thus improve the skills of its workforce, get access to technology and graduate to higher-value-added work – and not simply remain the low-cost supplier of labour. It is imperative to maintain this upward direction, since there are always poorer countries vying to take on the lower-value work.

---

[124] Mokyr, J., Vickers, C. and Ziebarth, N. L. (2015), 'The history of technological anxiety and the future of economic growth: Is this time different?', *Journal of Economic Perspectives*, 29(3), pp. 31–50, http://dx.doi.org/10.1257/jep.29.3.31.
[125] Bramucci, A. (2016), *Offshoring, Employment and Wages*, Working Paper, No. 71/2016, Berlin: Institute for International Political Economy, www.ipe-berlin.org/fileadmin/downloads/Papers_and_Presentations/IPE_WP_71.pdf.
[126] Acemoglu, D. and Restrepo, P. (2017), *Robots and Jobs: Evidence from US Labor Markets*, National Bureau of Economic Research Working Paper, No. 23285, http://www.nber.org/papers/w23285. Figures adjusted per thousand workers come from: Miller, C. C. (2017), 'Evidence That Robots Are Winning the Race for American Jobs', *New York Times*, 28 March 2017, https://www.nytimes.com/2017/03/28/upshot/evidence-that-robots-are-winning-the-race-for-american-jobs.html.

This process describes the path taken by Japan in the early 20th century, and subsequently emulated by the Asian high-growth economies of Hong Kong, Taiwan, Singapore and South Korea in the latter half of the century. It is known as the 'flying geese' model of development, because when the process is plotted on a graph, the series of swooning and drooping lines (that represent increasing and decreasing production over time, as the country enters and exits certain industries) resemble birds flying away.

Whereas rich countries have workers with more skills who may be able to adapt to the changes, workers in poor countries have yet to develop this resilience

But AI may mean that particular goose is cooked. Machines will replace labour for many jobs in developing countries, just as they will in advanced economies. This means the same pressure on employment and wages will hit developing countries, stalling the development model. Yet whereas rich countries have workers with more skills who may be able to adapt to the changes, workers in poor countries have yet to develop this resilience.

The development model has, moreover, relied on exports from countries with low-cost labour to richer ones. However, if the cost of manufacturing declines through automation (such as the robotic welder in car-making, to return to the example earlier in this chapter), a process of 'in-shoring' or 're-shoring' may happen, whereby work previously sent abroad comes back. More domestic manufacturing means fewer jobs in export industries in poor countries, and less global trade.[127]

As more work stays local, the use of migrant workers may decline in everything from construction (as in Dubai, for instance) to fruit-picking (as in California). If a host country's reliance on migrant labour is reduced through automation, one effect may be that remittances from overseas workers to their home countries fall, deepening those poorer countries' economic difficulties. An alternative scenario may be that if emerging markets are hit hard by AI, this may spur more migration to advanced countries. This would put even more downward pressure on unskilled wages in rich countries – and foster anti-foreigner sentiment and political tensions.

This suggests that AI may exacerbate inequalities among countries. Whoever owns the robots owns the means of production – to adapt Karl Marx to the era of AI. Yet just as rich countries may grow richer even as workers suffer lower wages, poorer countries will find it harder to find a way on to the development ladder. AI could 'erode the comparative advantage of much of the developing world,' warns David Autor, an economist at MIT.[128]

A less gloomy outlook is also possible. Tyler Cowen, an economist at George Mason University, takes it as a given that wealthy countries will automate manufacturing and produce more finished goods at home, rather than buy in from poor, export-oriented countries.[129] But he argues that 'trickle-down' growth is possible. Cowen makes the case that poor countries adopt the automation technology themselves from the West and forgo manufactured exports in favour of producing for domestic demand. Newer

[127] One potential benefit in this scenario is to the environment: less pollution is generated because of fewer goods being transported long distances.
[128] Autor quoted in Standage, T. (2016), 'The return of the machinery question', *The Economist*, 25 June 2016, https://www.economist.com/news/special-report/21700761-after-many-false-starts-artificial-intelligence-has-taken-will-it-cause-mass.
[129] Cowen, T. (2016), 'Economic development in an 'Average is over' world', George Mason University Working Paper.

Instead of being stuck making flat-panel screens for rich countries, a developing economy might use robots to lower costs and make TVs for local consumers – who then can get jobs making local content to fill the channels

and better jobs may then crop up around these products. So, instead of being stuck making flat-panel screens for rich countries, a developing economy might use robots to lower costs and make TVs for local consumers – who then can get jobs making local content to fill the channels.

There are advantages and drawbacks with this model. As AI technology dramatically lowers the cost of goods and services, the benefit is that items become much more affordable and people's lives are improved. The fast spread of inexpensive mobile phones in Africa is a notable example of this.

The drawback is that without manufacturing, the growth model will likely produce wide economic inequalities. This is because manufacturing-based growth requires investments in public education and infrastructure, and those improvements touch a wide segment of the population, including other businesses not directly related to the industry for which the improvements were designed. If AI means that these broad-based investments are not made, society overall will suffer even as the elites involved in AI-related businesses flourish. The backlash against the geeks looks pronounced in San Francisco; just wait until it reaches cramped cities in developing economies.

A nascent 'geopolitics of AI' is becoming apparent. Kai-Fu Lee, the chief executive of Sinovation Ventures, a major Chinese venture capital firm, argues that the AI lead of China and the US is now so dominant that every other nation will end up being an 'economic dependent' of one or other of them to access the technology – in part because countries with large populations will be saddled with mass joblessness.[130]

Yet that view seems overly cynical. There is a good chance that AI will not substantially change the economic balance of power among nations. The asymmetries that currently exist will remain if countries adopt AI at roughly the same pace, which the case of mobile phone technology suggests is broadly possible. China may reclaim its position as a low-cost producer not through cheap workers but through its forte in bots. Africa and Latin America will continue to lag behind, not because of a shortfall in human capital or good governance but because of a paucity of data and algorithms compared with their abundance in the West.

In so far as there are divides in the global economy – in access to education, in access to the internet, in access to medicine – these will probably persist with advances in AI. However, despite an 'AI gap' with the West, developing countries will still be better off: AI is poised to improve standards of living, as technologies tend to do.

[130] Lee, K.-F. (2017), 'The Real Threat of Artificial Intelligence', *New York Times*, 24 June 2017, https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html.

Preparing for the future

The economics of AI swing in opposite directions: the promise of efficiency and the peril of unemployment. Yet the debate has been dominated by anxiety over jobs. During the 2016 US presidential election and in the first year of his administration, Donald Trump stoked the worries of middle-class Americans that immigrants and free trade were stealing work. Automation has not thus far figured on his list of culprits, even though some economists would place it high up on theirs.

If the politics of economic disenfranchisement helped a populist gain power before the AI party has really even begun, just imagine the difficulties the world will experience when the economic impact of AI hits for real

If the politics of economic disenfranchisement helped a populist gain power before the AI party has really even begun, just imagine the difficulties the world will experience when the economic impact of AI hits for real. It can be expected that AI will be lambasted and used as a pretext to introduce protectionist policies if it is blamed for unemployment – even if there is a surge in output. Global trade may suffer. Domestically, AI will be branded as spurring 'jobless growth' that only benefits a handful of computer engineers, to the detriment of the struggling baristas, dog-walkers and yoga instructors who minister to them.

What can countries do to prepare? Three general and complementary ideas have been advanced: education, welfare and redistribution.

As regards education, there are calls for policies to encourage more targeted education, particularly in science, technology, engineering and maths. This is one of the main recommendations of a White House report on AI issued in the final weeks of the Obama administration.[131] It is not quite clear how this will truly stave off a jobs apocalypse. The traditional subjects are also the areas in which AI is making rapid advances. So there is a case to be made for rethinking education around social and caring skills.

There is also a substantial lag time: today's youngest students may not enter the labour force for perhaps two decades. Nevertheless, the idea is that more education cannot hurt – especially because research shows that learning in one area better equips people to acquire skills in others. This may help to overcome the skills mismatch and need for lifelong learning, which seems to be a bigger problem with AI than with other technologies.

In terms of welfare reforms, many have urged Nordic-style policies that provide benefits to individuals regardless of work status, in return for labour market liberalization that gives business freedom to manoeuvre. The model is termed 'flexicurity' (flexible security), and is broadly popular with individuals because it removes the sting of changing jobs or leaving the workforce temporarily. Business also supports it, since it means a more fluid labour market than if benefits are tied to the employer.[132] If AI means constant churn in the job market, flexicurity may strike a good social balance.

On redistribution policies to reduce inequality, the most ambitious calls are for a universal basic income. The idea is to redistribute economic gains (in this case from AI) more evenly throughout society. Although the idea predates AI, it has acquired more momentum with the rise of wealthy digital businesses and in light of the potential productivity boom that AI might unleash. Small trials have been taking

[131] National Science and Technology Council (2016), *Preparing For The Future of Artificial Intelligence, Automation, and the Economy*, Washington, D.C.: Office of Science and Technology Policy, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.
[132] Colin, N. and Palier, B. (2015), 'The Next Safety Net: Social Policy for a Digital Age', *Foreign Affairs*, 16 June 2015, https://www.foreignaffairs.com/articles/2015-06-16/next-safety-net.

place in Finland and Canada since 2017. However, critics argue that, on principle, provision of universal basic income removes the incentive to work, and, in practice, is unaffordable for state coffers.[133]

A more inventive policy to reduce inequality has been advanced by Jerry Kaplan, an AI expert at Stanford University and a Silicon Valley entrepreneur. He argues that government should offer generous tax breaks to companies based on the breadth of their share ownership, with the aim of encouraging the widest participation.[134] By this model, more people benefit from the fruits of AI than just the geeks who build it. Furthermore, since most people lack the capital to invest, he contends that governments should let people choose where they invest their national pension contributions while they are still young, so they can share in AI's spoils.

The advent of AI presents numerous other public policy and ethical challenges. The institution-building has begun. Well-funded foundations and academic centres have opened for business at Cambridge, Oxford, Harvard, MIT and elsewhere to bring together experts from different disciplines. An array of independent non-profit organizations are emerging as well.

Ultimately, AI will reshape the economy and society like so many other general-purpose technologies before it, from printing to electricity to computing. A parallel world is being created in which everything under the sun constantly generates and collects data; all things will continually improve their performance based on this information. Robots and algorithms will exceed human cognitive and physical performance, just as the pulley, the lever and the wheel enabled people to go beyond their muscular limits many millennia ago. Aristotle's lyre will finally play itself. We may find ourselves better off for it in some respects, but we are certainly not prepared for it.

[133] The Economist (2016), 'Basically flawed: Proponents of a basic income underestimate how disruptive it would be', *The Economist*, 4 June 2016, http://www.economist.com/news/leaders/21699907-proponents-basic-income-underestimate-how-disruptive-it-would-be-basically-flawed.
[134] Kaplan (2015), *Humans Need Not Apply*.

5. Conclusions and Recommendations

Hannah Bryce and Jacob Parakilas

Across most, if not all, sectors, the future of human endeavour will see more and more integration between humans and machines in both operational and decision-making roles. The driver of this change is – and will continue to be – the quest for greater efficiency, greater effectiveness and greater safety, all of which stand to be significantly improved through advances in AI. The challenge for policymakers, then, lies not in building the technology but rather in nurturing the governance frameworks needed to manage and regulate this integrated approach. The task is complex because of the manifold issues arising from such fundamental changes to the way we work, and is further compounded by the speed at which technology is evolving.

Managing the transition well will allow society to absorb the shock that increased automation and autonomy will have on the workplace. While this report has focused on AI with particular reference to the future of warfare, human security, and the economy and jobs, there are almost limitless potential impacts on the drivers of international politics. Some drivers are easy to envision at this stage – the use of AI to supercharge disinformation or to influence political processes such as election campaigns, for example. Others, like changes to legal or public health systems, might have more subtle or indirect effects.

The question of accountability is particularly salient due to the complexity of AI systems and the difficulty of translating their 'thought' processes to non-specialist audiences – what has come to be known as the 'black box' problem

The use of 'centaurs', in principle, could create a mechanism by which decision-making processes are enhanced but ultimate responsibility still lies in human hands. This is important when it comes to governance as it allows for accountability, a key tenet of liberal democracy. The question of accountability is particularly salient due to the complexity of AI systems and the difficulty of translating their 'thought' processes to non-specialist audiences – what has come to be known as the 'black box' problem. While work is being done on addressing this issue, AI processes and decisions remain largely opaque, especially to policymakers.[135] As AI further permeates people's lives, the process of assigning responsibility for harm caused by decisions made by (or with) algorithms can be extremely complicated, especially in situations where the impact of these decisions is not immediate or direct. Throughout, efforts to ensure responsibility and liability must be balanced with the risk of stifling innovation or potential, representing a new dimension of old challenges for policymakers.[136]

Questions of accountability are particularly pertinent when considering the military applications of AI. In 2017 prominent technologists, including Elon Musk of Tesla and Mustafa Suleyman of DeepMind, published an open letter in which they urged the UN to find a way to protect society from the potentially negative developments and uses of lethal autonomous weapons systems.[137] Indeed, the question of meaningful human control has been at the heart of discussions over how to legally and ethically control lethal autonomous weapons, including at the April 2018 UN Group of Governmental Experts meeting.

Moving towards a ban could be a useful strategy for policymakers at this time, even if its precise contours remain contentious. A dominant focus on such 'killer robots' – to use the term adopted by campaign groups and the media alike – risks obfuscating a more nuanced discussion of the much broader, non-lethal applications of AI.

---

[135] Knight, W. (2017), 'The Dark Secret at the Heart of AI', *MIT Technology Review*, 11 April 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.
[136] For one example of the complexity of these challenges, see Sweeney, L. (2013), 'Discrimination in Online Ad Delivery', SSRN, http://dx.doi.org/10.2139/ssrn.2208240.
[137] Gibbs, S. (2017), 'Elon Musk leads 116 experts calling for outright ban of killer robots', *Guardian*, 20 August 2017, https://www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war.

Removing from the equation the fundamental question of whether a robot should have the authority to decide whether or not to take a life might create a space for a much-needed public discussion around other areas and applications of AI.

As Missy Cummings observes, though, the critical issue is that much of the underlying research and technology that would drive and enable these weapons is being developed in the private sector, beyond the purview of state regulation, so a ban – though it would deplete the market – would likely not prevent the capacity being taken forward. Indeed, the question of who aside from states would be purchasing the technology gives rise to other concerns.

The 'Fourth Industrial Revolution' ushered in by AI will by some predictions create a range of new jobs that have not yet been imagined, while others see it destabilizing and undermining the current world order to detrimental effect

The availability of AI technology raises other concerns, related to the affordability and accessibility of technology as it advances, and the implications for societies that do not have the capacity or the resources to develop their own AI systems, as Heather Roff points out. There is a risk that AI will deepen already pronounced inequalities between advanced and less developed economies. While some countries may benefit indirectly from advances in AI – such as through the benefits to the humanitarian sector that AI may bring – there may be significant repercussions for societies whose workforce is replaced or streamlined by its application. The 'Fourth Industrial Revolution' ushered in by AI will, as Kenn Cukier sets out, by some predictions create a range of new jobs that have not yet been imagined, while others see it destabilizing and undermining the current world order to detrimental effect. These concerns underlie efforts – such as the May 2018 Toronto Declaration on protecting rights to equality and non-discrimination in machine learning systems – intended to avoid the deployment of AI systems designed without due regard for universal human rights.

Whichever predictions turn out to be true, managing the resulting changes to society will be integral to defining our subsequent relationship with AI. The challenge for policymakers is to grapple with the speed, scale and breadth at which AI can operate, and at which government, as a general rule, cannot. Failure to do so could mean that while the impact of AI will be international in scope, it may be divisive at a national level as governments struggle to get to grips with the range of possibilities that this technology offers.

Policy recommendations

Where AI is discussed in reference to public policy, the narrative tends to veer towards the extreme. Elon Musk's widely reported warning, in August 2017, that AI represents 'vastly more risk' than North Korea is a case in point.[138] While AI undoubtedly poses some significant risks that must be mitigated, discussion of these often crowds out sober analysis of the ways in which machine-aided decision-making is likely to change international politics in the relatively near term. Such analysis should have both near- and far-term goals. For the near term, the aim should be to achieve meaningful and measureable progress towards demystifying the technology and enabling productive conversations between those developing it and those who will be responsible for implementing and regulating it. At the same time, discussions around AI should not lose sight of the fundamental ways in which it may change the nature of international politics and power structures, and should aim to build up ethical and legal frameworks to manage those changes.

[138] David, J. E. (2017), 'Elon Musk issues a stark warning about A.I., calls it a bigger threat than North Korea', CNBC, 11 August 2017, https://www.cnbc.com/2017/08/11/elon-musk-issues-a-stark-warning-about-a-i-calls-it-a-bigger-threat-than-north-korea.html.

While no one can predict the exact trajectory that AI will take over the coming decades, it is clear that it will have an increasing and profound impact on society. To prepare for this transformation, this report makes a number of recommendations for policymakers:

- **AI expertise must not reside in only a small number of countries – or solely within narrow segments of the population.** As AI is entrusted with more and more significant responsibilities, programmers and policymakers alike must be more aware of its potential impact on existing structural inequalities. The processes by which AI systems are developed and deployed must be as inclusive as possible in order to mitigate societal risks, and inherent bias issues, at the point of inception. Policymakers must invest in developing home-grown talent and expertise in AI if countries are to be independent of the current dominant AI expertise that is concentrated particularly in China and the US. This will mean investing in education at all levels, both to foster a pipeline of leading AI engineers, and to ensure a workforce that is developing skills that AI possibly will not be able to replicate, such as those that value emotional capital.

- **Corporations, governments and foundations alike should allocate funding to develop and deploy AI systems with humanitarian goals.** There are significant advantages that AI could bring to the humanitarian sector, including through the use of complex datasets and planning algorithms that could, for instance, improve response times in humanitarian emergencies. More research needs to be done with discrete sectors to consider the specific implications that advances in AI will bring to them. Because AI development for humanitarian purposes is unlikely to be immediately profitable for the private sector, a concerted effort needs to be made to develop such systems on a not-for-profit basis.

- **Understanding of the capacities and limitations of artificially intelligent systems must not be the exclusive preserve of technical experts.** The information technology revolution will undoubtedly need more STEM graduates, but AI developers also need more 'soft' skills both at the operator level and in terms of integrating AI successfully into larger decision-making networks. Better education and training on what AI is – and, critically, what it is not – should be made as broadly available as possible.

- **Developing strong working relationships, particularly in the defence sector, between public and private AI developers is critical, as much of the innovation is taking place in the commercial sector.** The defence sector needs to find a way to harness and utilize the innovation that the rapidly evolving commercial market for AI technology is developing with, in some cases, greater resources to dedicate to it. Ensuring that intelligent systems charged with critical tasks can carry them out safely and ethically will require openness between different types of institutions.

- **Given the broad applicability of the technology, clear codes of practice are necessary to ensure that the benefits of AI can be shared widely while its concurrent risks are well managed.** Neither engineers nor policymakers alone possess the tools necessary to design, test and implement these codes – rather, they will require sustained and cooperative engagement between those communities. Policymakers and technologists should, moreover, understand the ways in which regulating artificially intelligent systems may be fundamentally different from regulating arms or trade flows, while also drawing relevant lessons from those models.

- **Particular attention must be paid by developers and regulators to the question of human–machine interfaces.** Artificial and human intelligence are fundamentally different, and interfaces between the two must be designed carefully, and reviewed constantly, in order to avoid misunderstandings that in many applications could have serious consequences.

It is essential for the public debate to move beyond an apocalyptic vision of robotic disruption on the one hand and a fanciful, automated idyll on the other. This is not a conversation about the future: AI is already in everyday use in mundane and not very spectacular ways – in areas such as navigation, text translation and retail. As the integration of AI applications with daily life continues, it is important that governments and publics alike understand what this means for now and for the future. Enabling an informed, nuanced and in-depth discussion will mean moving one step closer to that understanding.

About the Authors and Acknowledgments

Professor Mary L. 'Missy' Cummings is the director of the Humans and Autonomy Laboratory at Duke University. She received her bachelor's degree in mathematics from the US Naval Academy in 1988, her master's in space systems engineering from the Naval Postgraduate School in 1994, and her PhD in systems engineering from the University of Virginia in 2004. A naval officer and military pilot from 1988–99, she was one of the US Navy's first female fighter pilots. She is currently a professor in the Duke University Electrical and Computer Engineering and Computer Science departments. She is also a member of the US Department of Transportation's advisory committee on autonomous transportation, a fellow of the American Institute for Aeronautics and Astronautics, and a member of the board of directors of Veoneer, Inc. Her research interests include human supervisory control, human–unmanned vehicle interaction, human–autonomous systems collaboration, human–robot interaction, human systems engineering, and the ethical and social impact of technology.

Author's acknowledgments
The author would like to thank the many sponsors of her research that helped to inform this work, including the National Science Foundation, the Office of Naval Research, NASA, and the US Army and Air Force.

Dr Heather Roff is currently a senior research fellow in the Department of Politics and International Relations at the University of Oxford, a research scientist in the Global Security Initiative at Arizona State University, and has held faculty positions at the Korbel School of International Studies at the University of Denver, the University of Waterloo, and the United States Air Force Academy. She is also an associate research fellow at the Leverhulme Centre for the Future of Intelligence at the University of Cambridge, a research fellow at New America in the Cybersecurity Initiative and the Future of War Project. Her research interests include the law, policy and ethics of emerging military technologies, such as autonomous weapons, artificial intelligence, robotics and cyber, as well as international security and human rights protection.

Author's acknowledgments
The author would like to thank the many sponsors of her research that helped to inform this piece, including the Future of Life Foundation, the government of Canada, and Google DeepMind.

Kenneth Cukier is a senior editor at *The Economist*, and host of its weekly podcast on technology. He is the co-author, with Viktor Mayer-Schönberger, of *Big Data: A Revolution That Transforms How We Live, Work, and Think* (2013). He is an associate fellow at Saïd Business School, University of Oxford, researching artificial intelligence. Previously, he was a foreign correspondent for two decades in Europe, Asia and America. He was a research fellow at Harvard's Kennedy School of Government in 2002–04. He is a board director of The Open String, a member of the Chatham House Council, and a member of the Council on Foreign Relations. He is a frequent commentator in the media and keynote speaker on AI.

Author's acknowledgments
The author would like to thank his colleagues at *The Economist*, whose coverage and conversation nurtured many of the ideas in his chapter for this report. Thanks are also due to the University of Oxford's Saïd Business School, where he is an associate fellow.

Dr Jacob Parakilas is the deputy head of the US and the Americas Programme at Chatham House. He previously worked as head of weapons at Action on Armed Violence, a London-based NGO working on armed violence reduction worldwide. His academic research has covered nuclear weapons, violent non-state actors in the Middle East, and drug violence and the arms trade in Mexico. He has also worked at the World Security Institute, the Arms Control Association and the US Department of Homeland Security.

Hannah Bryce was formerly assistant head of the International Security Department (ISD) at Chatham House, and worked on the secretariat of the Global Commission on Internet Governance. Her areas of research included human security (particularly related to explosive weapons), the role of women in the military, cybersecurity and its policy implications, and the impact and effectiveness of the humanitarian sector, particularly the UN. She is co-editor of the *Journal of Cyber Policy*.

Authors' acknowledgments
The authors would like to thank the many peer reviewers who reviewed parts and all of this report, and whose expertise and comments have informed, enhanced and helped shape this final version. They also thank the Chatham House Director's Research and Innovation Fund, which provided the basis for funding the report. The authors are grateful to all the Chatham House staff who have supported the development and production of this report, including Henry Dodd, Patricia Lewis, Jo Maher, Kitty McKendrick, Courtney Rice, Mike Tsang and Xenia Wickett.

About the International Security Department

The International Security Department (ISD) at Chatham House produces leading research for policymakers and opinion-shapers. The department's permanent team of specialist security experts and extensive network of associate fellows produce research and foster debate on a wide range of security issues and their impact on the international arena, including:

- Cybersecurity – Critical infrastructure protection, cyberwarfare and internet governance;
- Human Security – Arms control, biosecurity, and drugs and organized crime; and
- Weapons and technology – Nuclear disarmament and non-proliferation, chemical weapons prevention, and drones and the future of warfare.

ISD's research is accessible to both the public and private sectors, and seeks to develop discussions on how to tackle current and emerging threats or challenges to security. Through publications, events and international engagement, the department draws on a range of subject matters within Chatham House and through its international partnerships. ISD adopts a multidisciplinary approach to security, injecting diverse perspectives and regional insights into today's debates.

About the US and the Americas Programme

Building on the independent, international reputation of Chatham House, the US and the Americas Programme provides a unique external perspective on US foreign affairs.

The work of the programme aims to:

- Develop a contextual understanding of the transformations taking place within the US and internationally, to analyse how they affect US foreign policymaking;
- Offer insights on America's likely future international direction;
- Influence responses from key allies and others to US policy decisions; and
- Highlight to US policymakers the impact of their decisions on the international system.

The US and the Americas Programme commissions and conducts original research, hosts discussions and offers specialist commentary on the changing role of the US in the world.

The programme's in-house subject specialists and its international network of associate fellows together provide in-depth expertise in both geopolitical and thematic areas.

Independent thinking since 1920

The Royal Institute of International Affairs
Chatham House
10 St James's Square, London SW1Y 4LE
T +44 (0)20 7957 5700 F +44 (0)20 7957 5710
contact@chathamhouse.org www.chathamhouse.org

Charity Registration Number: 208223

ISBN 978-1-78413-212-5