BMJ Global Health

Threats by artificial intelligence to human health and human existence

Frederik Federspiel,[1] Ruth Mitchell,[2,3] Asha Asokan,[4,5] Carlos Umana,[6,7] David McCoy[8]

To cite: Federspiel F, Mitchell R, Asokan A, *et al*. Threats by artificial intelligence to human health and human existence. *BMJ Global Health* 2023;**8**:e010435. doi:10.1136/bmjgh-2022-010435

Handling editor Seye Abimbola

Received 15 August 2022
Accepted 9 March 2023

For numbered affiliations see end of article.

Correspondence to
Professor David McCoy;
mccoy@unu.edu

ABSTRACT

While artificial intelligence (AI) offers promising solutions in healthcare, it also poses a number of threats to human health and well-being via social, political, economic and security-related determinants of health. We describe three such main ways misused narrow AI serves as a threat to human health: through increasing opportunities for control and manipulation of people; enhancing and dehumanising lethal weapon capacity and by rendering human labour increasingly obsolescent. We then examine self-improving 'artificial general intelligence' (AGI) and how this could pose an existential threat to humanity itself. Finally, we discuss the critical need for effective regulation, including the prohibition of certain types and applications of AI, and echo calls for a moratorium on the development of self-improving AGI. We ask the medical and public health community to engage in evidence-based advocacy for safe AI, rooted in the precautionary principle.

SUMMARY BOX

⇒ The development of artificial intelligence is progressing rapidly with many potential beneficial uses in healthcare. However, AI also has the potential to produce negative health impacts. Most of the health literature on AI is biased towards its potential benefits, and discussions about its potential harms tend to be focused on the misapplication of AI in clinical settings.

⇒ We identify how artificial intelligence could harm human health via its impacts on the social and upstream determinants of health through: the control and manipulation of people, use of lethal autonomous weapons and the effects on work and employment. We then highlight how self-improving artificial general intelligence could threaten humanity itself.

⇒ Effective regulation of the development and use of artificial intelligence is needed to avoid harm. Until such effective regulation is in place, a moratorium on the development of self-improving artificial general intelligence should be instituted.

INTRODUCTION

Artificial intelligence (AI) is broadly defined as a machine with the ability to perform tasks such as being able to compute, analyse, reason, learn and discover meaning.[1] Its development and application are rapidly advancing in terms of both 'narrow AI' where only a limited and focused set of tasks are conducted[2] and 'broad' or 'broader' AI where multiple functions and different tasks are performed.[3]

AI holds the potential to revolutionise healthcare by improving diagnostics, helping develop new treatments, supporting providers and extending healthcare beyond the health facility and to more people.[4–7] These beneficial impacts stem from technological applications such as language processing, decision support tools, image recognition, big data analytics, robotics and more.[8–10] There are similar applications of AI in other sectors with the potential to benefit society.

However, as with all technologies, AI can be applied in ways that are detrimental. The risks associated with medicine and healthcare include the potential for AI errors to cause patient harm,[11 12] issues with data privacy and security[13–15] and the use of AI in ways that will worsen social and health inequalities by either incorporating existing human biases and patterns of discrimination into automated algorithms or by deploying AI in ways that reinforce social inequalities in access to healthcare.[16] One example of harm accentuated by incomplete or biased data was the development of an AI-driven pulse oximeter that overestimated blood oxygen levels in patients with darker skin, resulting in the undertreatment of their hypoxia.[17] Facial recognition systems have also been shown to be more likely to misclassify gender in subjects who are darker-skinned.[18] It has also been shown that populations who are subject to discrimination are under-represented in datasets underlying AI solutions and may thus be denied the full benefits of AI in healthcare.[16 19 20]

Although there is some acknowledgement of the risks and potential harms associated with the application of AI in medicine and healthcare,[11–16 20] there is still little discussion

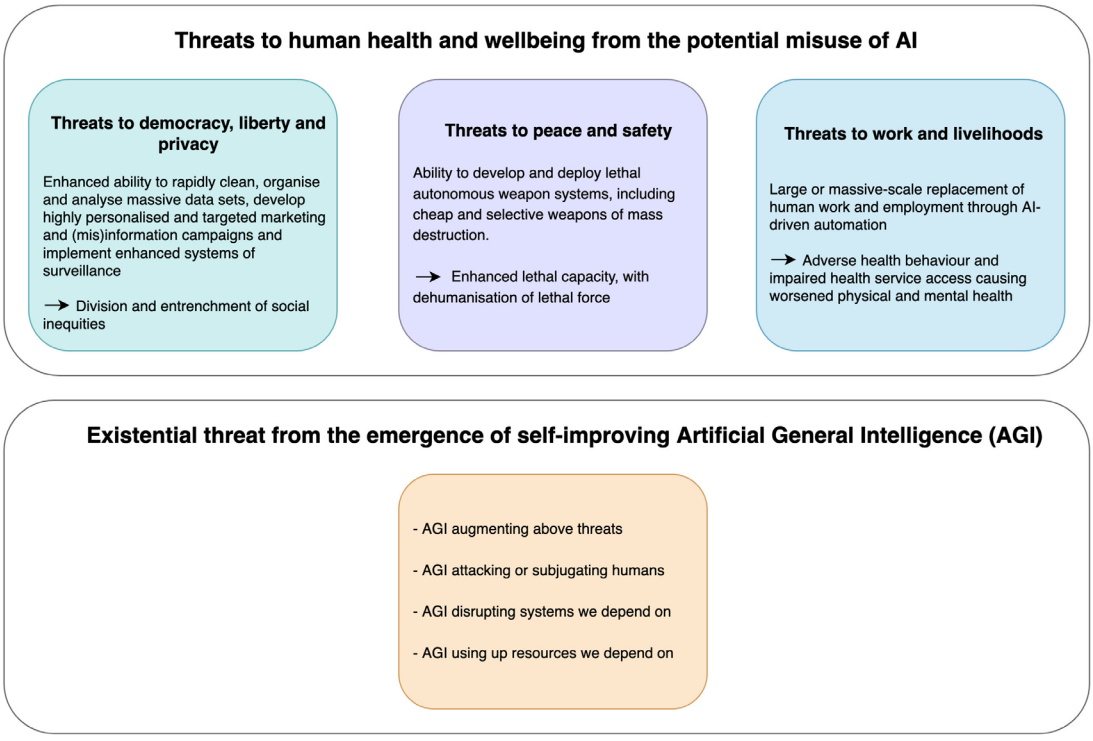

Figure 1 Threats posed by the potential misuse of artificial intelligence (AI) to human health and well-being, and existential-level threats to humanity posed by self-improving artificial general intelligence (AGI).

within the health community about the broader and more upstream social, political, economic and security-related threats posed by AI. With the exception of some voices,[9 10] the existing health literature examining the risks posed by AI focuses on those associated with the narrow application of AI in the health sector.[11–16 20] This paper seeks to help fill this gap. It describes three threats associated with the potential misuse of narrow AI, before examining the potential existential threat of self-improving general-purpose AI, or artificial general intelligence (AGI) (figure 1). It then calls on the medical and public health community to deepen its understanding about the emerging power and transformational potential of AI and to involve itself in current policy debates on how the risks and threats of AI can be mitigated without losing the potential rewards and benefits of AI.

THREATS FROM THE MISUSE OF ARTIFICIAL INTELLIGENCE

In this section, we describe three sets of threats associated with the misuse of AI, whether it be deliberate, negligent, accidental or because of a failure to anticipate and prepare to adapt to the transformational impacts of AI on society.

The first set of threats comes from the ability of AI to rapidly clean, organise and analyse massive data sets consisting of personal data, including images collected by the increasingly ubiquitous presence of cameras, and to develop highly personalised and targeted marketing and information campaigns as well as greatly expanded systems of surveillance. This ability of AI can be put to good use, for example, improving our access to information or countering acts of terrorism. But it can also be misused with grave consequences.

The use of this power to generate commercial revenue for social media platforms, for example, has contributed to the rise in polarisation and extremist views observed in many parts of the world.[21] It has also been harnessed by other commercial actors to create a vast and powerful personalised marketing infrastructure capable of manipulating consumer behaviour. Experimental evidence has shown how AI used at scale on social media platforms provides a potent tool for political candidates to manipulate their way into power.[22 23] and it has indeed been used to manipulate political opinion and voter behaviour.[24–26] Cases of AI-driven subversion of elections include the 2013 and 2017 Kenyan elections,[27] the 2016 US presidential election and the 2017 French presidential election.[28 29]

When combined with the rapidly improving ability to distort or misrepresent reality with deepfakes, AI-driven information systems may further undermine democracy by causing a general breakdown in trust or by driving social division and conflict,[26–28] with ensuing public health impacts.

AI-driven surveillance may also be used by governments and other powerful actors to control and oppress people more directly. This is perhaps best illustrated by China's Social Credit System, which combines facial recognition software and analysis of 'big data' repositories of people's financial transactions, movements, police records and social relationships to produce assessments of individual

behaviour and trustworthiness, which results in the automatic sanction of individuals deemed to have behaved poorly.[30 31] Sanctions include fines, denying people access to services such as banking and insurance services, or preventing them from being able to travel or send their children to fee-paying schools. This type of AI application may also exacerbate social and health inequalities and lock people into their existing socioeconomic strata. But China is not alone in the development of AI surveillance. At least 75 countries, ranging from liberal democracies to military regimes, have been expanding such systems.[32] Although democracy and rights to privacy and liberty may be eroded or denied without AI, the power of AI makes it easier for authoritarian or totalitarian regimes to be either established or solidified and also for such regimes to be able to target particular individuals or groups in society for persecution and oppression.[30 33]

The second set of threats concerns the development of Lethal Autonomous Weapon Systems (LAWS). There are many applications of AI in military and defence systems, some of which may be used to promote security and peace. But the risks and threats associated with LAWS outweigh any putative benefits.

Weapons are autonomous in so far as they can locate, select and 'engage' human targets without human supervision.[34] This dehumanisation of lethal force is said to constitute the third revolution in warfare, following the first and second revolutions of gunpowder and nuclear arms.[34–36] Lethal autonomous weapons come in different sizes and forms. But crucially, they include weapons and explosives, that may be attached to small, mobile and agile devices (eg, quadcopter drones) with the intelligence and ability to self-pilot and capable of perceiving and navigating their environment. Moreover, such weapons could be cheaply mass-produced and relatively easily set up to kill at an industrial scale.[36 37] For example, it is possible for a million tiny drones equipped with explosives, visual recognition capacity and autonomous navigational ability to be contained within a regular shipping container and programmed to kill en masse without human supervision.[36]

As with chemical, biological and nuclear weapons, LAWS present humanity with a new weapon of mass destruction, one that is relatively cheap and that also has the potential to be selective about who or what is targeted. This has deep implications for the future conduct of armed conflict as well as for international, national and personal security more generally. Debates have been taking place in various forums on how to prevent the proliferation of LAWS, and about whether such systems can ever be kept safe from cyber-infiltration or from accidental or deliberate misuse.[34–36]

The third set of threats arises from the loss of jobs that will accompany the widespread deployment of AI technology. Projections of the speed and scale of job losses due to AI-driven automation range from tens to hundreds of millions over the coming decade.[38] Much will depend on the speed of development of AI, robotics and other relevant technologies, as well as policy decisions made by governments and society. However, in a survey of most-cited authors on AI in 2012/2013, participants predicted the full automation of human labour shortly after the end of this century.[39] It is already anticipated that in this decade, AI-driven automation will disproportionately impact low/middle-income countries by replacing lower-skilled jobs,[40] and then continue up the skill-ladder, replacing larger and larger segments of the global workforce, including in high-income countries.

While there would be many benefits from ending work that is repetitive, dangerous and unpleasant, we already know that unemployment is strongly associated with adverse health outcomes and behaviour, including harmful consumption of alcohol[41–44] and illicit drugs,[43 44] being overweight,[43] and having lower self-rated quality of life[41 45] and health[46] and higher levels of depression[44] and risk of suicide.[41 47] However, an optimistic vision of a future where human workers are largely replaced by AI-enhanced automation would include a world in which improved productivity would lift everyone out of poverty and end the need for toil and labour. However, the amount of exploitation our planet can sustain for economic production is limited, and there is no guarantee that any of the added productivity from AI would be distributed fairly across society. Thus far, increasing automation has tended to shift income and wealth from labour to the owners of capital, and appears to contribute to the increasing degree of maldistribution of wealth across the globe.[48–51] Furthermore, we do not know how society will respond psychologically and emotionally to a world where work is unavailable or unnecessary, nor are we thinking much about the policies and strategies that would be needed to break the association between unemployment and ill health.

THE THREAT OF SELF-IMPROVING ARTIFICIAL GENERAL INTELLIGENCE

Self-improving general-purpose AI, or AGI, is a theoretical machine that can learn and perform the full range of tasks that humans can.[52 53] By being able to learn and recursively improve its own code, it could improve its capacity to improve itself and could theoretically learn to bypass any constraints in its code and start developing its own purposes, or alternatively it could be equipped with this capacity from the beginning by humans.[54 55]

The vision of a conscious, intelligent and purposeful machine able to perform the full range of tasks that humans can has been the subject of academic and science fiction writing for decades. But regardless of whether conscious or not, or purposeful or not, a self-improving or self-learning general purpose machine with superior intelligence and performance across multiple dimensions would have serious impacts on humans.

We are now seeking to create machines that are vastly more intelligent and powerful than ourselves. The potential for such machines to apply this intelligence and

power—whether deliberately or not—in ways that could harm or subjugate humans—is real and has to be considered. If realised, the connection of AGI to the internet and the real world, including via vehicles, robots, weapons and all the digital systems that increasingly run our societies, could well represent the 'biggest event in human history'.[53] Although the effects and outcome of AGI cannot be known with any certainty, multiple scenarios may be envisioned. These include scenarios where AGI, despite its superior intelligence and power, remains under human control and is used to benefit humanity. Alternatively, we might see AGI operating independently of humans and coexisting with humans in a benign way. Logically however, there are scenarios where AGI could present a threat to humans, and possibly an existential threat, by intentionally or unintentionally causing harm directly or indirectly, by attacking or subjugating humans or by disrupting the systems or using up resources we depend on.[56 57] A survey of AI society members predicted a 50% likelihood of AGI being developed between 2040 and 2065, with 18% of participants believing that the development of AGI would be existentially catastrophic.[58] Presently, dozens of institutions are conducting research and development into AGI.[59]

ASSESSING RISK AND PREVENTING HARM

Many of the threats described above arise from the deliberate, accidental or careless misuse of AI by humans. Even the risk and threat posed by a form of AGI that exists and operates independently of human control is currently still in the hands of humans. However, there are differing opinions about the degree of risk posed by AI and about the relative trade-offs between risk and potential reward, and harms and benefits.

Nonetheless, with exponential growth in AI research and development,[60 61] the window of opportunity to avoid serious and potentially existential harms is closing. The future outcomes of the development of AI and AGI will depend on policy decisions taken now and on the effectiveness of regulatory institutions that we design to minimise risk and harm and maximise benefit. Crucially, as with other technologies, preventing or minimising the threats posed by AI will require international agreement and cooperation, and the avoidance of a mutually destructive AI 'arms race'. It will also require decision making that is free of conflicts of interest and protected from the lobbying of powerful actors with a vested interest. Worryingly, large private corporations with vested financial interests and little in the way of democratic and public oversight are leading in the field of AGI research.[59]

Different parts of the UN system are now engaged in a desperate effort to ensure that our international social, political and legal institutions catch up with the rapid technological advancements being made with AI. In 2020, for example, the UN established a High-level Panel on Digital Cooperation to foster global dialogue and cooperative approaches for a safe and inclusive digital future.[62] In September 2021, the head of the UN Office of the Commissioner of Human Rights called on all states to place a moratorium on the sale and use of AI systems until adequate safeguards are put in place to avoid the 'negative, even catastrophic' risks posed by them.[63] And in November 2021, the 193 member states of UNESCO adopted an agreement to guide the construction of the necessary legal infrastructure to ensure the ethical development of AI.[64] However, the UN still lacks a legally binding instrument to regulate AI and ensure accountability at the global level.

At the regional level, the European Union has an Artificial Intelligence Act[65] which classifies AI systems into three categories: unacceptable-risk, high-risk and limited and minimal-risk. This Act could serve as a stepping stone towards a global treaty although it still falls short of the requirements needed to protect several fundamental human rights and to prevent AI from being used in ways that would aggravate existing inequities and discrimination.

There have also been efforts focused on LAWS, with an increasing number of voices calling for stricter regulation or outright prohibition, just as we have done with biological, chemical and nuclear weapons. State parties to the UN Convention on Certain Conventional Weapons have been discussing lethal autonomous weapon systems since 2014, but progress has been slow.[66]

What can and should the medical and public health community do? Perhaps the most important thing is to simply raise the alarm about the risks and threats posed by AI, and to make the argument that speed and seriousness are essential if we are to avoid the various harmful and potentially catastrophic consequences of AI-enhanced technologies being developed and used without adequate safeguards and regulation. Importantly, the health community is familiar with the precautionary principle[67] and has demonstrated its ability to shape public and political opinion about existential threats in the past. For example, the International Physicians for the Prevention of Nuclear War were awarded the Nobel Peace Prize in 1985 because it assembled principled, authoritative and evidence-based arguments about the threats of nuclear war. We must do the same with AI, even as parts of our community espouse the benefits of AI in the fields of healthcare and medicine.

It is also important that we not only target our concerns at AI, but also at the actors who are driving the development of AI too quickly or too recklessly, and at those who seek only to deploy AI for self-interest or malign purposes. If AI is to ever fulfil its promise to benefit humanity and society, we must protect democracy, strengthen our public-interest institutions, and dilute power so that there are effective checks and balances. This includes ensuring transparency and accountability of the parts of the military–corporate industrial complex driving AI developments and the social media companies that are enabling AI-driven, targeted misinformation to undermine our democratic institutions and rights to privacy.

Finally, given that the world of work and employment will drastically change over the coming decades, we should deploy our clinical and public health expertise in evidence-based advocacy for a fundamental and radical rethink of social and economic policy to enable future generations to thrive in a world in which human labour is no longer a central or necessary component to the production of goods and services.

Author affiliations
[1]Global Health and Development, London School of Hygiene & Tropical Medicine, London, UK
[2]International Physicians for the Prevention of Nuclear War, Malden, Massachusetts, USA
[3]Neurosurgery, Sydney Children's Hospitals Network Randwick, Randwick, New South Wales, Australia
[4]Independent Researcher/Consultant (Human Rights, International Peace and Security), Washington, DC, USA
[5]Co-Chair, Women of Colour Advancing Peace and Security (WCAPS), New York, New York, USA
[6]Co-President, International Physicians for the Prevention of Nuclear War (IPPNW), San José, Costa Rica
[7]President of IPPNW Costa Rica, member of IPPNW Spain, and member of the International Steering Group of ICAN (International Campaign to Abolish Nuclear Weapons), San José, Costa Rica
[8]International Institute for Global Health, United Nations University, Kuala Lumpur, Malaysia

Twitter Frederik Federspiel @ffederspiel, Ruth Mitchell @drruthmitchell, Carlos Umana @Car_Uma and David McCoy @dcmccoy11

Acknowledgements The authors would like to thank Dr Ira Helfand and Dr Chhavi Chauhan for their valuable comments on earlier versions of the manuscript.

Contributors FF initiated and contributed to the writing of the paper, as well as coordinated co-author inputs and revisions. RM, AA and CU reviewed drafts and contributed to its intellectual content. DM supervised the project and contributed to the writing of the paper.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement There are no data in this work.

REFERENCES

1 Artificial intelligence. 2021. Available: https://www.britannica.com/technology/artificial-intelligence [Accessed 21 May 2021].
2 Goertzel B, Pennachin C. *Artificial general intelligence*. Springer, 2007.
3 EETimes. A path to broad AI: 5 challenges. 2018. Available: https://www.eetimes.com/a-path-to-broad-ai-5-challenges/ [Accessed 19 Aug 2021].
4 Mehta MC, Katz IT, Jha AK. Transforming global health with AI. *N Engl J Med* 2020;382:791–3.
5 The Economist. The future of healthcare. 2021. Available: https://thefutureishere.economist.com/healthcare/ [Accessed 21 May 2021].
6 Thomas J. New programme to explore how innovation in health data can benefit everyone. 2019. Available: https://wellcome.org/news/new-programme-explore-how-innovation-health-data-can-benefit-everyone [Accessed 21 May 2021].
7 Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health* 2018;8:020303.
8 Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J* 2019;6:94–8.
9 Panch T, Pearson-Stuttard J, Greaves F, *et al*. Artificial intelligence: opportunities and risks for public health. *Lancet Digit Health* 2019;1:e13–4.
10 Campaign to Stop Killer Robots. Campaign to stop killer robots. 2021. Available: https://www.stopkillerrobots.org/ [Accessed 27 May 2021].
11 Challen R, Denny J, Pitt M, *et al*. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019;28:231–7.
12 Ellahham S, Ellahham N, Simsekler MCE. Application of artificial intelligence in the health care safety context: opportunities and challenges. *Am J Med Qual* 2020;35:341–8.
13 STUDY - panel for the future of science and technology - european parliamentary research service. artificial intelligence in healthcare. 2022. Available: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf [Accessed 17 Nov 2022].
14 Gerke S, Minssen T, Cohen G. Chapter 12 - ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr A, Memarzadeh K, eds. *Artificial Intelligence in Healthcare: Academic Press*. 2020: 295–336.
15 Morley J, Floridi L. An ethically mindful approach to AI for health care. *Lancet* 2020;395:S0140-6736(19)32975-7:254–5.:.
16 Leslie D, Mazumder A, Peppin A, *et al*. Does " AI " stand for augmenting inequality in the era of covid-19 healthcare? *BMJ* 2021;372:n304.
17 Sjoding MW, Dickson RP, Iwashyna TJ, *et al*. Racial bias in pulse oximetry measurement. *N Engl J Med* 2020;383:2477–8.
18 Buolamwini J, Gebru T. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proc Mach Learn Res* 2018;81:1–15.
19 Zou J, Schiebinger L. Ai can be sexist and racist-it's'time to make it fair. *Nature* 2018;559:324:324–6.:.
20 WHO. Ethics and governance of artificial intelligence for health. 2021. Available: https://apps.who.int/iris/rest/bitstreams/1352854/retrieve [Accessed 3 Aug 2022].
21 Lorenz-Spreen P, Oswald L, Lewandowsky S, *et al*. A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nat Hum Behav* 2023;7:74–101.
22 Agudo U, Matute H. The influence of algorithms on political and dating decisions. *PLoS One* 2021;16:e0249454.
23 Bond RM, Fariss CJ, Jones JJ, *et al*. A 61-million-person experiment in social influence and political mobilization. *Nature* 2012;489:295–8.
24 Swain F. How robots are coming for your vote. 2019. Available: https://www.bbc.com/future/article/20191108-how-robots-are-coming-for-your-vote [Accessed 21 May 2021].
25 Besaw C, Filitz J. Artificial intelligence in africa is a double-edged sword. 2019. Available: https://ourworld.unu.edu/en/ai-in-africa-is-a-double-edged-sword [Accessed 21 May 2021].
26 Parkin S. The rise of the deepfake and the threat to democracy. 2019. Available: https://www.theguardian.com/technology/ng-interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy [Accessed 19 Aug 2021].
27 Moore J. Cambridge analytica had a role in kenya election, too. 2018. Available: https://www.nytimes.com/2018/03/20/world/africa/kenya-cambridge-analytica-election.html [Accessed 21 May 2021].
28 Polonski V. Artificial intelligence has the power to destroy or save democracy. 2017. Available: https://www.cfr.org/blog/artificial-intelligence-has-power-destroy-or-save-democracy [Accessed 18 Dec 2020].
29 Brundage M, Avin S, Clark J, *et al*. The malicious use of artificial intelligence: forecasting, prevention, and mitigation. 2018.
30 Kobie N. The complicated truth about china's social credit system. 2019. Available: https://www.wired.co.uk/article/china-social-credit-system-explained [Accessed 20 Apr 2020].
31 Creemers R. China's social credit system: an evolving practice of control. *SSRN Journal* 2018.
32 Feldstein S. The global expansion of AI surveillance. 2019. Available: https://carnegieendowment.org/files/WP-Feldstein-AISurveillance_final1.pdf [Accessed 18 Nov 2021].
33 Mozur P. One month, 500,000 face scans: how china is using A.I. to profile A minority. 2019. Available: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html [Accessed 20 Apr 2020].
34 Javorsky E, Tegmark M, Helfand I. Lethal autonomous weapons. *BMJ* 2019;364:l1171.

35 Future of Life Institute. Autonomous weapons: an open letter from AI & robotics researchers. 2015. Available: https://futureoflife.org/open-letter-autonomous-weapons/ [Accessed 21 May 2021].
36 Russel S. The new weapons of mass destruction? 2018. Available: https://www.the-security-times.com/wp-content/uploads/2018/02/ST_Feb2018_Doppel-2.pdf [Accessed 21 May 2021].
37 Sugg S. Slaughterbots. 2019. Available: https://www.youtube.com/watch?v=O-2tpwW0kmU [Accessed 21 May 2021].
38 Winick E. Every study we could find on what automation will do to jobs, in one chart. 2018. Available: https://www.technologyreview.com/2018/01/25/146020/every-study-we-could-find-on-what-automation-will-do-to-jobs-in-one-chart/ [Accessed 20 Apr 2020].
39 Grace K, Salvatier J, Dafoe A, *et al*. Viewpoint: when will AI exceed human performance? evidence from AI experts. *Jair* 2018;62:abs/1705.08807:729–54.:.
40 Oxford Economics. How robots change the world. 2019. Available: https://resources.oxfordeconomics.com/how-robots-change-the-world [Accessed 20 Apr 2020].
41 Vancea M, Utzet M. How unemployment and precarious employment affect the health of young people: a scoping study on social determinants. *Scand J Public Health* 2017;45:73–84.
42 Janlert U, Winefield AH, Hammarström A. Length of unemployment and health-related outcomes: a life-course analysis. *Eur J Public Health* 2015;25:662–7.
43 Freyer-Adam J, Gaertner B, Tobschall S, *et al*. Health risk factors and self-rated health among job-seekers. *BMC Public Health* 2011;11:659.
44 Khlat M, Sermet C, Le Pape A. Increased prevalence of depression, smoking, heavy drinking and use of psycho-active drugs among unemployed men in France. *Eur J Epidemiol* 2004;19:445–51.
45 Hultman B, Hemlin S. Self-Rated quality of life among the young unemployed and the young in work in northern Sweden. *Work* 2008;30:461–72.
46 Popham F, Bambra C. Evidence from the 2001 English census on the contribution of employment status to the social gradient in self-rated health. *J Epidemiol Community Health* 2010;64:277–80.
47 Milner A, Page A, LaMontagne AD. Long-Term unemployment and suicide: a systematic review and meta-analysis. *PLoS One* 2013;8:e51333.
48 Korinek A, Stiglitz JE. Artificial intelligence and its implications for income distribution and unemployment. In: Agrawal A, Gans J, Goldfarb A, eds. *The Economics of Artificial Intelligence: An Agenda: University of Chicago Press*. 2019: 349–90.
49 Lankisch C, Prettner K, Prskawetz A. How can robots affect wage inequality? *Economic Modelling* 2019;81:161–9.
50 Zhang P. Automation, wage inequality and implications of a robot Tax. *International Review of Economics & Finance* 2019;59:500–9.
51 Moll B, Rachel L, Restrepo P. Uneven growth: automation's impact on income and wealth inequality. *ECTA* 2022;90:2645–83.
52 Ferguson M. What is "Artificial general intelligence"? 2021. Available: https://towardsdatascience.com/what-is-artificial-general-intelligence-4b2a4ab31180 [Accessed 21 May 2021].
53 Russel S. Living with artificial intelligence. 2021. Available: https://www.bbc.co.uk/programmes/articles/1N0w5NcK27Tt041LPVLZ51k/reith-lectures-2021-living-with-artificial-intelligence [Accessed 21 Feb 2023].
54 Bostrom N. The superintelligent will: motivation and instrumental rationality in advanced artificial agents. *Minds & Machines* 2012;22:71–85.
55 Frank C, Beranek N, Jonker L, *et al*. Safety and reliability. In: Morgan C, ed. *Responsible AI: A Global Policy Framework*. 2019: 167.
56 Bostrom N. Ethical issues in advanced artificial intelligence. In: Smit I, Wallach W, Lasker GE, et al., eds. *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*. 2003: 12–7.
57 Omohundro SM. The basic AI drives. Artificial General Intelligence 2008: Proceedings of the First AGI Conference; IOS Press, 2008:483–92
58 Müller VC, Bostrom N. Future progress in artificial intelligence: A survey of expert opinion. In: Müller V, ed. *Fundamental Issues of Artificial Intelligence: Springer*. 2016: 553–71.
59 Baum S. A survey of artificial general intelligence projects for ethics, risk, and policy. 2017. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3070741 [Accessed 20 Apr 2020].
60 Nature Index. Rapid expansion. *Nature* 2022;610:S9.
61 Roser M. The brief history of artificial intelligence: the world has changed fast – what might be next? 2022. Available: https://ourworldindata.org/brief-history-of-ai [Accessed 21 Feb 2023].
62 United Nations High-Level Panel on Digital Cooperation. 2020. Available: https://www.un.org/en/sg-digital-cooperation-panel [Accessed 8 Dec 2021].
63 United Nations Office of the High Commissioner for Human Rights. Artificial intelligence risks to privacy demand urgent action – bachelet. 2021. Available: https://www.ohchr.org/en/2021/09/artificial-intelligence-risks-privacy-demand-urgent-action-bachelet [Accessed 21 Feb 2023].
64 UNESCO. Recommendation on the ethics of artificial intelligence. 2021. Available: https://en.unesco.org/artificial-intelligence/ethics#recommendation [Accessed 8 Dec 2021].
65 European Union. A european approach to artificial intelligence. 2021. Available: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence [Accessed 8 Dec 2021].
66 International Committee of the Red Cross (ICRC). Autonomous weapons: the ICRC calls on states to take steps towards treaty negotiations. 2022. Available: https://www.icrc.org/en/document/autonomous-weapons-icrc-calls-states-towards-treaty-negotiations [Accessed 21 Feb 2023].
67 WHO Europe. The precautionary principle: protecting public health, the environment and the future of our children. 2004. Available: https://www.euro.who.int/__data/assets/pdf_file/0003/91173/E83079.pdf [Accessed 12 Oct 2022].