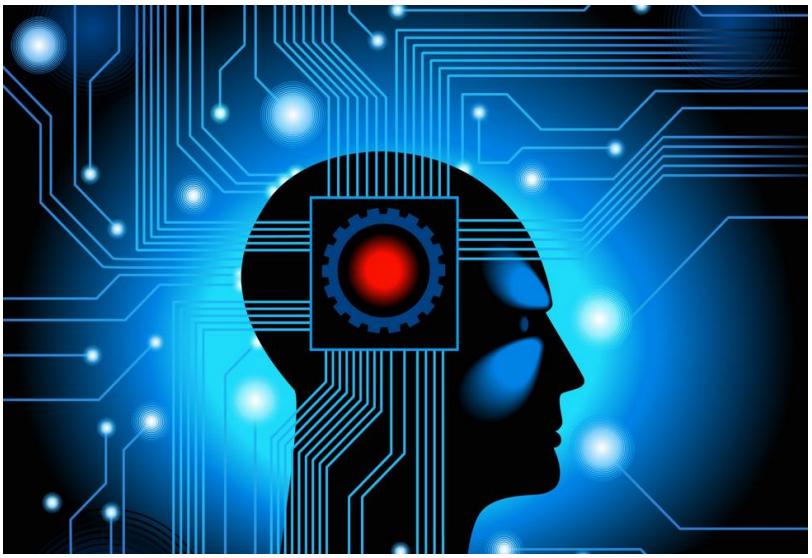
Georgetown Journal of International Affairs

The Security Implications and Existential Crossroads of Artificial Intelligence

By Nayef Al-Rodhan



- wired.co.uk

Emerging technologies and their possible implications for ethics, security, and even human existence have increasingly gained ground in the past two decades. Some innovations have resulted in obvious security and existential threats: a world with nuclear arms, for example. The potential of other technological shifts, however, has been more mixed. Biotechnologies, genetic engineering, and stems cells have given rise to controversial debates in which advocacy groups on both sides have convincingly put forward pros and cons. The Internet has revolutionized everything from markets to family communication in ways both beneficial and harmful. The age of artificial Intelligence (AI) has shown itself to be similarly Janus-like in its potential to alter our lives both positively and negatively. On the one hand, AI has demonstrated its usefulness in predictive speech and typing software, robotics, and unmanned aircraft technology. On the other, these and many other AI-enabled platforms raise profound concerns about oversight.

AI is also unique among emergent technologies because it can learn and evolve without human input. This fact alone demands a policy approach that recognizes not only the immediate implications of AI itself but also what might happen because of the potential range of resultant technologies. In short, AI poses challenges for security and policymaking not merely of magnitude but of precedent. Further, AI forces us to consider our relationship with technology in ways that were never previously relevant—including the possibility of entering into competition with, and even being superseded by, our own creations.

The advent of AI brings with it numerous implications for the futures of global security, conflicts, and human dignity. The extensive use of drones, both for military and commercial purposes, is a rightly controversial current debate. But the uses of AI in unmanned aircraft

are mere glimmers of what is to come. In the later stages of the industrial revolution, industrialization in factories rendered some jobs previously performed by human beings obsolete. AI appears to portend the inevitable complete removal of human beings from combat scenarios in numerous military-strategic areas.

AI applications facilitate real-time adaptation to contingencies without requiring the presence of people on the ground. Unmanned drones, for instance, are used to provide continuous surveillance and small robots are deployed in missions to counter improvised explosive devices. U.S. Army researchers are now working to develop intelligent robots that can successfully navigate in different environments by following voice commands and instructions by a human. Furthermore, the U.S. Defense Advanced Research Projects Agency (DARPA) launched an AI program in 2013 to help integrate machine-learning capacity in a wide variety of military weapons. Other teams of scientists are now exploring ways to create robots with a moral compasses and in-built senses of right or wrong that have the ability to pick the ethical course of action on the battlefield.

Two immediate consequences of this transition to battlefield AI are especially noteworthy. The first reflects the relative ease of convincing the public or another decision-making body to engage in violent conflict in cases where the use of AI technology assures minimal human casualties. Given that President Obama's strategy to "degrade and ultimately destroy" the Islamic State in Iraq and Syria (ISIS), for example, attempted to explicitly avoid committing further on-the-ground American troops, wars that do not involve risk of bodily harm to soldiers continue to be much easier sells to both the public and to government bodies. These assurances are potentially problematic not only because they tend to work against even the most circumspect evaluation of a war's justness, but also because they encourage a point of view that underestimates the destabilizing effect of all military engagements, regardless of battlefield casualties. This point of view often overlooks warfare's terrible track record of noncombatant casualties and harm to nonmilitary parties. The history of recorded warfare demonstrates that far more civilians than soldiers have died as a result of military engagements, a trend that has significantly worsened in the era of modern technology. This fact alone should evidence a need for additional reflection about the part AI will play in the future of warfare.

A related area of concern is the role of judgment regarding entry into and conduct during interstate conflict (*jus ad bellum and jus in bello*). Any AI machine expected to make decisions in war should pass some variation of the Turing test, which was devised by British mathematician Alan Turing in 1950 to assess whether a particular machine exhibits intelligence equivalent to or beyond that of a human. But the worry is that a robotic soldier or a sufficiently sophisticated AI drone could easily pass a version of the Turing test and yet utterly fail to uphold *jus in bello*'s fundamental commitment to non-combatant immunity, or *jus ad bellum*'s supposed principal of non-aggression. Therefore, if AI is to play a role in military engagement, this potential must be closely monitored and constrained by international norms.

Second, as I have previously argued, a heavy reliance on AI machines would create further inequalities in war because of the unequal availability of such technologies to certain countries. This will make the outcome of interstate conflict far more directly a matter of superior technology and which nations or peoples have the resources to attain it. This availability gap could serve to exacerbate and reinforce preexisting global inequalities. This could also conceivably result in asymmetric battlefield casualties where countries that have access to AI technology will suffer fewer human losses compared to those countries that do not. Other questions about AI's use and application are relevant too. Could conscious machines be sensitive to human welfare? Could they replicate the human motivation to cooperate in order to avoid the "state of nature," which Hobbes defined as a state of a perpetual war and lack of effective higher authority to arbitrate disputes? How can we expect

robots to understand, relate to, and execute the basic norms of social cooperation and political order?

Beyond its potential military applications, the nature and use of AI should also be monitored and regulated in non-combat settings. AI has achieved an almost ubiquitous presence in our everyday lives in the machines and applications we use in the workplace, at home, and beyond. Learning software, like the popular "Swipe" texting key—an app that learns user's tendency to use particular words and phrases and becomes predictive of what a user is trying to say or is about to say next—is an example of the sort of AI that is coming to play a significant role in everyday life. A similar technology, developed by Intel, is responsible for the speech-assistance software used by British physicist Stephen Hawking, whose degenerative ALS rendered him unable to speak unassisted by machinery in 1985. Nevertheless, while acknowledging the benefit he receives from AI, Hawking has vocalized concerns that complete AI could bring about the end of the human race. With the capacity to learn and improve at near-limitless rates, full AIs would quickly become superior to human beings, constrained as we are by long and slow evolutionary processes.

While the dystopian vision of runaway or out-of-control AI still appears like something out of science fiction, today's rate of technological innovation serves as a reminder that we may be headed in that direction. The collective of hackers and activists known as Anonymous has demonstrated the fearsome capacity of AI programs even at their current stage of development: at the outset of the Arab Spring in 2011, leading members of the group clogged the networks of Tunisia's governing regime. Within 24 hours, the websites of the president, prime minister, and that of the Tunisian stock exchange had been brought down. Simple AI can learn to avoid spam filters, avoid fraud detection, and disguise itself as various different forms of online protocol. And these features are minimal compared to the more advanced capabilities to which AI might lead—the ability of a fully AI machine to make strategic decisions about which governments to isolate or which weapons systems to activate, for instance.

Regardless of how close to or far from the realization of such capabilities we are, the fact that the possibility exists in principle should motivate dialogue and careful control over the development of AI. Alongside environmental degradation and large-scale human rights violations, artificial intelligence represents yet another critical challenge that requires interstate collaboration and the shoring up of international law to preserve the safety and dignity of human beings in both our contemporary and future world.

Nayef Al-Rodhan *is a philosopher, neuroscientist and geostrategist. He is an Honorary Fellow of St. Antony's College, University of Oxford, UK, and Senior Fellow and Director of the Centre for the Geopolitics of Globalization and Transnational Security at the Geneva Centre for Security Policy, Geneva, Switzerland. He is the author of Sustainable History and the Dignity of Man: A Philosophy of History and Civilisational Triumph (Zürich: LIT VERLAG GmbH & Co. KG Wien, 2009).*

This article was originally published in the Georgetown Journal of International Affairs on April 2, 2015.
http://journal.georgetown.edu/the-security-implications-and-existential-crossroads-of-artificial-intelligence/