An International Journal in Philosophy, Religion, Politics, and the Arts

Existenz

Volume 17, No 1, Spring 2022 ISSN 1932-1066

Risks of Weaponry Integrated with Artificial Intelligence

Gregory M. Reichberg
Peace Research Institute Oslo, Norway
greg.reichberg@prio.org

Abstract: Assessing the risks associated with automated weapon platforms is a major theme in Robert Latiff's book, *Future Peace*. While the development of artificial intelligence has been an aspect of weapon design over the last three decades, the newest forms of this technology that are based on machine learning, have considerably raised the stakes, leading to heated debates regarding lethal autonomous weapon systems and algorithmic warfare. With good reason, Latiff points out that the high-velocity of weapon systems that are integrated with this technology renders their effects increasingly difficult to manage for the service-personnel responsible for targeting decisions in combat settings. Safety must be thought anew in the age of AI, especially when new weapon systems are rolled out in an international context in which states are actively pursuing technological superiority over their peers.

Keywords: Latiff, Robert H.; arms control; artificial intelligence; machine learning; risk; safety; weaponry.

At the outset, I would like to say how much I have learned from engaging with Robert Latiff's excellent book.[1] It covers a wide range of issues (as befits its title), all of which he has treated cohesively and lucidly. First, I will review the points that most resonated with me, and afterwards I will consider some points for which a fuller explanation would be beneficial.

Not having read very deeply into the general literature on ethics and technology, I much enjoyed (but was also troubled by) the section "Fascination with Technology and Superweapons" in chapter 2, that is titled "Urges to Violence." Given my European perspective, Latiff's concern seemed to me as being the result of a distinctively American attitude; I find that other cultures appear to be more reticent in this regard. Even cultures that show a strong affection for technology, such as, for example, in Japan with its widespread embrace of robotics, this affinity does not spill over to weapons *per se*. Several years ago, Scott Sagan conducted a survey regarding United States perspectives toward the use of nuclear weapons. It showed surprisingly low support for the idea of a nuclear taboo. Most of those polled thought that a nuclear bomb could rightly be used in a foreign war, if doing so would protect the lives of United States combatants when facing even "a nonnuclear-armed adversary."[2] By contrast, it is safe to assume that currently some European countries would hesitate to propose the acquisition of autonomous weapon systems because of the public outrage this

[1] Robert H. Latiff, *Future Peace: Technology, Aggression, and the Rush to War*, Notre Dame, IN: University of Notre Dame Press, 2022. [Henceforth cited as *FP*]

[2] Scott D. Sagan and Benjamin A. Valentino, "Revisiting Hiroshima in Iran: What Americans Really Think about Using Nuclear Weapons and Killing Noncombatants," *International Security* 42/1 (Summer 2017), 41-79, here p. 45.

would likely provoke. This is visible in a recent UK government document on military use of artificial intelligence,[3] where the term "Lethal Autonomous Weapons Systems (LAWS)" appears only once and other, less contentious terms, such as "machine-speed command and control," "Human-Machine Teaming," or "autonomous...combat vehicles" are employed instead. This strategic use of terminology is likely motivated by a desire to steer public opinion away from contentious debates that would undermine the proposed policy to reorient the country's military strategy toward accelerated integration of artificial intelligence weapon technologies.

I agree with Latiff when he emphasizes the importance of keeping the focus on safety in international arms control discussions, especially as pertains to cyber and autonomous weapon systems, and I benefitted from his discussion to this effect in chapter 4 on "Avoiding War." Safety is an issue around which consensus could be built, as all countries have an investment in protecting their military personnel and citizenry from unintended harm, and to prevent the accidental inception of war. Former United States Secretary of Defense Richard Danzig authored a report, aptly titled *Technology Roulette*, where he explains how cybersecurity and artificial intelligence applications bear a resemblance to biological weapons insofar as they can easily be released into the digital ecosystem where much harm can be done to friends and foes alike.[4] Similar to what was accomplished in the Nuclear Non-Proliferation Treaty of 1968, these emerging technologies also stand in need of regulation by way of bilateral or multilateral agreements. While at the present juncture it is highly unlikely that Russia would take part in negotiations toward such agreements, China, for instance, has shown an interest in this topic. One should not discount the significance of dialogue with countries such as India, Israel, and Turkey, all of which are investing heavily in military applications that utilize artificial intelligence. Mitt Regan and Jovana Davidovic have advocated for the importance of promoting transparency regarding safety requirements for these enhanced weapon systems. This can be done by publishing criteria for the testing, evaluation, verification, and validation (TEVV) of these systems. Regan and Davidovich argue that the publication of these safety requirements beneficially signals "to other states that they will not be disadvantaged by likewise committing to use AI-enabled weapons only after such review."[5]

The inter-state discussions on autonomous weapons systems by the Group of Governmental Experts (GGE) at the Conference on Certain Conventional Weapons (CCW) that have been underway in Geneva since 2016 have largely broken down. Active creation of new such forums is imperative. Here I might venture that the discussions in Geneva have been overly narrow due to their concentration on machine autonomy. However, as Latiff shows convincingly throughout his book, the risks inherent to artificial intelligence applications arise also when these technologies are designed to function merely in support capacities, for example in order to issue targeting recommendations (when the actual decisions will be taken by human personnel). In such situations, humans will continue to give final orders, but when decisions are made under severe time pressure, reliance on machine-generated information can lead to disastrous outcomes. My point is that international arms control discussions must encompass not only autonomous systems (where human operators cede targeting decisions to the machine) but must also include command-and-control (C2) systems that support human decision-making, where safety issues often arise, as has been well-noted by Merel Ekelhof.[6] Decision-support systems are already widely in use (hence the safety concerns are already relevant) in contrast to systems based on deep learning that are mainly still under

[3] United Kingdom Ministry of Defence, "Defence Artificial Intelligence Strategy," 15 June 2022 https://www.gov.uk/government/publications/defence-artificial-intelligence-strategy/defence-artificial-intelligence-strategy.

[4] Richard Danzig, *Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority*, Washington, DC: Center for A New American Security, June 2018.

[5] Mitt Regan and Jovana Davidovic, "Just Preparation for War and AI-Enabled Weapons," *Frontiers in Big Data* 6 (12 May 2023), 1-6, here p. 5. [Henceforth cited as *JPW*]

[6] Merel A.C. Ekelhof, "Lifting the fog of war: Autonomous weapons and human control through the lens of targeting," *Naval War College Review* 71/3 (Summer 2018), 61-94; https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=5125&context=nwc-review.

development. Moreover, alongside C2 it is highly pertinent that space-based weaponry and weapon support systems are also brought into the discussion regarding arms control. Latiff raises an important point when he observes how the Outer Space Treaty of 1967 requires updating in light of the new weapon technologies that have been developed over the last decades. His succinct depiction of the risks posed by these technologies, and the need for energetic diplomacy to reduce these risks, are valuable points raised within *Future Peace*.

Regarding Latiff's discussion of risk issues concerning command-and-control systems, I think that the analysis would have benefitted from a more nuanced differentiation of the various technologies that underlie machine automation. For instance, older forms of automation, such as the ones implemented in the Aegis Combat System, and which are still operative today, function mainly by way of knowledge-based systems that represent early forms of artificial intelligence. This method of utilizing artificial intelligence is very well understood, and systems such as the Aegis have been exposed to rigorous TEVV (Testing, Evaluation, Validation, Verification) assessments throughout the years and, subsequently, the training protocols have been refined so that the benefits of their use for protection against missile attacks outweigh the known risks.

In his depiction of risks, I come away with the impression that Latiff is tarring the entire endeavor of using artificial intelligence for enhancing military readiness, when in fact most of his reservations seem to apply to the newer forms of automation that lean heavily on the variant of machine learning known as "neural networks" or "deep learning." It is true that proposals have been made (and some weapon systems have been developed) that use deep learning for automated target recognition and engagement; its underlying technology (for instance, computer vision) is comparable to civilian uses of it, such as driverless cars or surveillance systems. It is apropos of these technologies that the author has good reason to signal the risks. The United States Department of Defense seems to be cognizant of this as well (perhaps more so than Latiff acknowledges) as reflected most recently in the new version of DoD Directive 3000.09, which mandates a stringent review process for systems developed with autonomous or semi-autonomous functions.[7] This comes on the heels of intense discussions in the United States Armed Forces and NATO regarding the theme of "responsible AI."

My question to Latiff is, hence, whether he thinks that these discussions, as well as the ethical and procurement guidelines that have been developed, will inevitably fail to ensure reduction of risks, given the very nature of the technologies involved? In other words, does he believe that this is a Sisyphean task, which would mean in consequence that the best course of action is simply to ban machine learning the military's targeting protocols? A related question concerns the trade-off between benefits and risks. Clearly, arguments can be made that the military advantages of using artificial intelligence operations, enhanced ISR (Intelligence, Surveillance and Reconnaissance), precision targeting, increased protection of personnel, reduced harm to civilians and property might outweigh the likely downsides, provided that the enemy forces do not have equal or superior technological capabilities.

I do however agree with the main thrust of Latiff's conclusion that the eventual risks will be mitigated not only by thorough testing of the attendant software but also by a serious investment in training programs for the human personnel who will deploy these weapon platforms.

Another topic that caught my attention is Latiff's claim, made at several junctures in the book, that the use of advanced technologies such as artificial intelligence systems will make outbreaks of armed conflict more likely. Latiff's reasoning seems to be twofold. On the one hand, he argues that when autonomous weapon systems are deployed military personnel will be removed from the dangers of the battlefield. With this reduction of risk to one's own troops, the political leadership will worry less about the adverse public effects of resorting to war, and the threshold to engaging in warfare will be lowered. Hence, according to Latiff risk-free war (for the side initiating armed force) will lead to more frequent wars and the world will be worse off as a result. On the other hand, Latiff notes how the use of new technologies that are not well understood or tested (for example, the

[7] Office of the Under Secretary of Defense for Policy, *DoD Directive 3000.09: Autonomy in Weapon Systems*, 25 January 2023, https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

so-called black box phenomenon where the outcomes proposed by artificial intelligence computations are untraceable before and after deployment, even by the system designers) could lead to miscalculations or misapprehensions or misinterpretations that might accidentally trigger the outbreak of war.

The two reasonings (the one based on risk-free war and the other on technology-induced error) seem intuitively plausible to me. However, intuition alone is never enough to ground a conclusion. In addition, some empirical evidence should also be provided, or at least realistic counter-examples or opposing arguments need to be explored. Relating to the first line of reasoning, namely the attempt to achieve risk free war, Stephanie Carvin maintains that this is a predominantly American approach to warfighting that emerged from the hubris that followed the First Gulf War and the attendant claims regarding, what had been dubbed in the 1990s, a revolution in military affairs.[8] This hubris accounts for the US proclivity to engage in wars far afield from home and its proximate security interests. On this reading, it is not the technologies themselves that lower the threshold to war, but the American fascination with these technologies that brings about this result. In the hands of a different military force, perhaps a Dutch, Norwegian, Japanese, or Indian force, this first line of reasoning might lead to a different conclusion.

Approaching the same argument from yet another perspective, Nathan Leys adopts the analytical framework of international security studies (in the strategic deterrence spirit of Thomas Schelling) in order to assess whether the use of autonomous weapon systems would necessarily increase the occurrence of wars.[9] Entertaining a negative reply to this question, Leys writes:

> Autonomy could afford the United States an off-ramp by providing a plausible cover: the potentially accidental nature of the violation of an ally's sovereignty means a military response is neither legally required nor morally warranted. In short, AWS could provide a face-saving alternative for leaders trying to de-escalate a crisis...In an interesting twist on the debate about whom to hold responsible in the event of an AWS's malfunction, the most life-saving answer in a crisis may be no one: If there is no one to blame, there is no one to bomb. [*AWS* 60]
>
> Additionally, because nonautonomous weapons like stealth bombers and remote-control UAVs can already carry out retaliatory strikes without significant operational risk to US soldiers, autonomy per se is not likely to be a unique reason why negligible operational risk means the United States might choose to escalate in such crises. [*AWS* 61]

Similarly, should an unmanned aircraft be shot down, there would probably be less incentive to retaliate than it would be in a situation in which a human combatant had died:

> A failure to respond to the death of a military member would prove politically disastrous for a US leader, but destroyed AWS do not have grieving families. [*AWS* 61]

In quoting Leys, my point is not that he demonstrates conclusively how new digital technologies are unlikely to make wars more frequent (he also entertains the countervailing arguments), but rather to indicate how one should be cautious on this question: More research on the *ad bellum* risks of the new weapon technologies is sorely needed and should be the precondition for advancing conclusive arguments on both sides.

Although not expressly mentioned by Latiff, an additional basis for concern that AI enabled weaponry might lead to increased rates of warfare arises apropos of what is commonly termed "the security dilemma." Such a dilemma arises when State A fears that the growing strength of State B's military capability will soon give State B a decisive military advantage. Consequently, to improve its odds of military success, State A attacks State B preemptively, leading to a war that might not otherwise have occurred. It is conceivable that new weaponry that is integrated with artificial intelligence may indeed shift the balance of power between states. Warning against this trend, Vladimir Putin is quoted by the Associated Press to have said in 2017 that "the one who becomes the leader in this sphere will be ruler of the world," to which he added the caveat that "it would be strongly undesirable if someone wins a monopolist position."[10]

[8] Stephanie Carvin, "How Not to War", *International Affairs* 98/5 (September 2022), 1695-1716.

[9] Nathan Leys, "Autonomous Weapon Systems and International Crises," *Strategic Studies Quarterly* 12/1 (Spring 2018), 48-73. Henceforth cited as *AWS*]

[10] Associated Press News, "Putin: Leader in Artificial Intelligence will Rule the World," 1 September 2017, https://apnews.com/article/bb5628f2a7424a10b3e38b07f4eb90d4.

When compared to slower human-operated, remote-control systems, Regan and Davidovic caution that

> the nature of AI-enabled weapons may intensify a security dilemma because of the perceived decisive advantage of operating at machine speed. [*JPW* 4]

Even if this advanced weaponry does not lead to the outbreak of war, worries about an emerging arms race, loss of geopolitical power, and the risk of military defeat, could lead states to field AI-enabled weaponry that has not been adequately tested (short-circuiting the needed TEVV process which, for advanced military systems can take upwards of eight years).[11] The proliferation of unsafe weapon systems creates unsafe conditions that may result in accidental firings, conflict escalation, retaliation. Latiff's cautionary words about the grave dangers of high-speed warfare dovetail closely with these reflections on the security dilemma and complement them well.

In conclusion, I would emphasize that despite the obvious risks, it is highly unlikely that states will abandon these new weapon technologies that occupy an increasingly important place in military planning. In this respect, initiatives to promote a ban on lethal autonomous weapon systems appear increasingly quixotic. More productive, in my opinion, would be to focus on ways of regulating, both at the national and international levels, the development and use of these systems. This would require a renewed and sustained attention to safety, as well as diplomatic efforts in reaching a consensus between states regarding limits for military uses of artificial intelligence. Ideally, states might agree to exclude computational command functions from specific applications, such as nuclear command and control, that is, in cases in which the downside of machine error would be unacceptably large.

[11] Personal communication with the CEO of a leading Norwegian defense company, dated 9 December 2022.