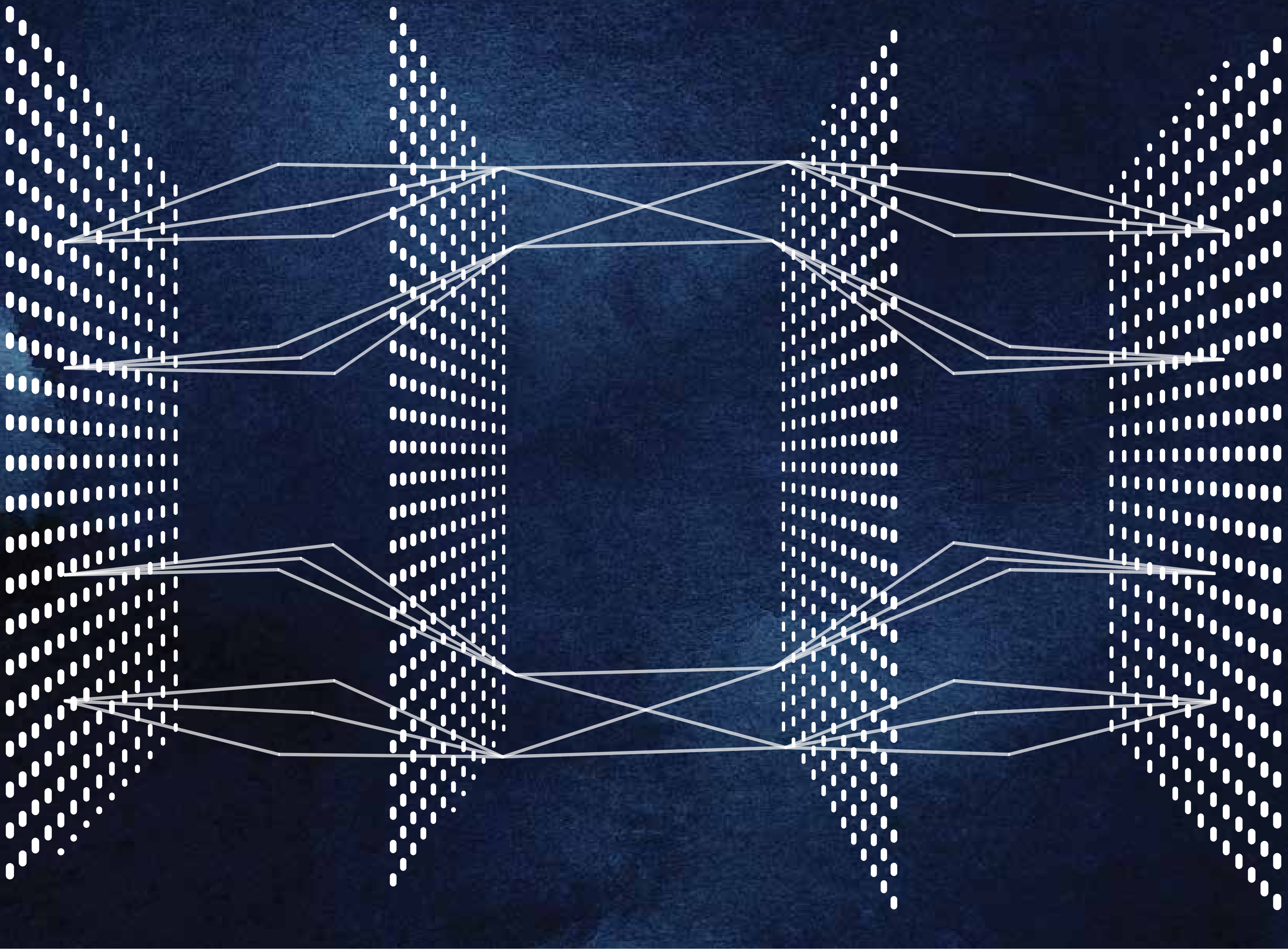
The Militarization of Artificial Intelligence

August 2019 | United Nations, New York, NY



Stanley Center FOR PEACE AND SECURITY

UNODA United Nations Office for Disarmament Affairs

STIMSON

Foreword

We are pleased to present to you this workshop summary and the associated discussion papers on The Militarization of Artificial Intelligence.

In his agenda for disarmament, *Securing Our Common Future, United Nations* Secretary-General António Guterres stated that, "Arms control has always been motivated by the need to keep ahead of the challenges to peace and security raised by science and technology" and emerging means and methods of warfare.

While revolutionary technologies hold much promise for humanity, when taken up for military uses they can pose risks for international peace and security. The challenge is to build understanding among stakeholders about a technology and develop responsive solutions to mitigate such risks.

That is where we might be today with military applications of artificial intelligence (AI).

There can be little doubt that AI has potential uses that could improve the health and well-being of individuals, communities, and states, and help meet the UN'Sustainable Development Goals. However, certain uses of AI could undermine international peace and security if they raise safety concerns, accelerate conflicts, or loosen human control over the means of war.

These papers emerge from a series of discussions coconvened by the UN Office for Disarmament Affairs, the Stanley Center for Peace and Security, and the Stimson Center. It was made possible through a generous contribution by the government of Switzerland. The organizers are particularly indebted to Reto Wollenmann and Beatrice Müller of the Swiss Department of Foreign Affairs for their thought leadership and guidance throughout the project. We are grateful to Jennifer Spindel, Paul Scharre, Vadim Kozyulin, and colleagues at the China Arms Control and Disarmament Association for their discussion papers. We also thank those experts who participated in the workshop for their thoughtful presentations and contributions.

A unique feature of this project was its multistakeholder composition, acknowledging the growing importance, in particular, of tech firms to security discussions. We hope this provides a starting point for more robust dialogues not just among governments but also industry and research institutions, as stakeholders endeavor to maximize the benefits of AI while mitigating the misapplication of this important technology.

Brian Finlay | President and CEO, Stimson Center
Benjamin Loehrke | Program Officer, Stanley Center for Peace and Security
Chris King | Senior Political Affairs Officer, UN Office of Disarmament Affairs

Workshop Summary

Multistakeholder Perspectives on the Potential Benefits, Risks, and Governance Options for Military Applications of Artificial Intelligence

Melanie Sisson | Defense Strategy and Planning Program, Stimson Center

Few developments in science and technology hold as much promise for the future of humanity as the suite of computer-science-enabled capabilities that falls under the umbrella of artificial intelligence (AI). AI has the potential to contribute to the health and well-being of individuals, communities, and states, as well as to aid fulfillment of the United Nations' 2030 agenda for Sustainable Development Goals. As with past revolutionary technologies, however, AI applications could affect international peace and security, especially through their integration into the tools and systems of national militaries.

In recognition of this, UN Secretary-General António Guterres, in his agenda for disarmament, *Securing Our Common Future*, stresses the need for UN member states to better understand the nature and implications of new and emerging technologies with potential military applications and the need to maintain human control over weapons systems. He emphasizes that dialogue among governments, civil society, and the private sector is an increasingly necessary complement to existing intergovernmental processes.

Such an approach is particularly relevant for AI, which, as an enabling technology, is likely to be integrated into a broad array of military applications but is largely being developed by private sector entities or academic institutions for different, mostly civilian, purposes.

To facilitate a conversation between disparate stakeholders on this topic, the UN Office for Disarmament Affairs, the Stimson Center, and the Stanley Center for Peace and Security convened an initial dialogue on the intersection of AI and national military capabilities. Over two days at UN headquarters in New York, experts from member states, industry, academia, and research institutions participated in a workshop on The Militarization of Artificial Intelligence.

Discussion within the workshop was candid and revealed that the implications for international peace and security of AI's integration into national militaries remains to a large extent unclear. Consequently, uncertainty about the domains in which and the purposes for which AI will be used by national militaries poses practical challenges to the design of governance mechanisms. This uncertainty generates fear and heightens perceptions of risk. These dynamics reflect the early stage of discourse on military applications of AI and reinforce the need for active and consistent engagement.

Workshop participants acknowledged and were mindful of the need for precision when referring to the large body of tools compressed into the term "AI," most notably by distinguishing between machine-assisted decision making and machine autonomy. The result was a rich discussion that identified three topical areas in need of ongoing learning and dialogue among member states and other stakeholders:

- **Potential Risks of Military Applications of AI:** There undoubtedly are risks posed by applications of AI within the military domain; it is important, however, to not be alarmist in addressing these potential challenges.

- **Potential Benefits of Military Application of AI:** There is a need to consider more fully the potential positive applications of AI within the military domain and to develop state-level and multilateral means of capturing these benefits safely.

– **Potential Governance of Military Applications of AI:** There are considerable challenges to international governance posed by these emergent technologies, and the primary work of stakeholders will be to devise constructs that balance the tradeoffs made between innovation, capturing the positive effects of AI, and mitigating or eliminating the risks of military AI.

Potential Risks of Military Applications of Artificial Intelligence

The risks of introducing artificial intelligence into national militaries are not small. Lethal autonomous weapon systems (LAWS) receive popular attention because such systems are easily imagined and raise important security, legal, philosophical, and ethical questions. Workshop participants, however, identified multiple other risks from military applications of AI that pose challenges to international peace and security.

Militaries are likely to use AI to assist with decision making. This may be through providing information to humans as they make decisions, or even by taking over the entire execution of decision-making processes. This may happen, for example, in communications-denied environments or in environments such as cyberspace, in which action happens at speeds beyond human cognition. While this may improve a human operator's or commander's ability to exercise direct command and control over military systems, it could also have the opposite effect. AI affords the construction of complex systems that can be difficult to understand, creating problems of transparency and of knowing whether the system is performing as expected or intended. Where transparency is sufficiently prioritized in AI design, this concern can be reduced. Where it is not, it becomes possible that errors in AI systems will go unseen—whether such errors are accidental or caused deliberately by outside parties using techniques like hacking or data poisoning.

Participants debated whether AI can be used effectively to hack, distort, or corrupt the functions of command-and-control structures, including early warning systems for nuclear weapons. Specific note was made, however, that the integration of multiple AI-enabled systems could make it harder to identify command-and-control malfunctions. Such integration is a likely direction for advancement in military applications of AI.

Participants also discussed how advances in AI interact with human trust in the machine-based systems they use. Increasing complexity could make AI systems harder to understand and, therefore, encourage the use of trust rather than transparency. Increased trust means that errors and failures are even less likely to be detected.

The concern was also expressed that the desire for—or fear of another's—decision-making speed may contribute to acting quickly on information aggregated and presented by AI. This pressure can increase the likelihood that decision makers will be prone to known automation biases, including rejection of contradictory or surprising information. So too might the addition of speed create pressures that work against caution and deliberation, with leaders fearing the consequences of delay. Speed can be especially destabilizing in combat, where increases in pace ultimately could surpass the human ability to understand, process, and act on information. This mismatch between AI speed and cognition could degrade human control over events and increase the destructiveness of violent conflict.

Although participants worry about the potential for lone actors to use AI-enabled tools, these concerns are moderated by their inability to apply them at large scale. More problematic to participants is the potential for national-level arms racing. The potential ill effects of AI arms racing are threefold. First, arms-race dynamics have in the past led to high levels of government spending that were poorly prioritized and inefficient. Second, arms racing can generate an insecurity spiral, with actors perceiving others' pursuit of new capabilities as threatening. Third, the development of AI tools for use by national militaries is in a discovery phase, with government and industry alike working to find areas for useful application. Competition at the industry and state levels might, therefore, incentivize fast deployment of new and potentially insufficiently tested capabilities, as well as hiding of national AI priorities and progress. These characteristics of arms racing—high rates of investment, a lack of transparency, mutual suspicion and fear, and a perceived incentive to deploy first—heighten the risk of avoidable or accidental conflict.

Potential Benefits of Military Applications of Artificial Intelligence

For national militaries, AI has broad potential beyond weapons systems. Often referred to as a tool for jobs that are "dull, dirty, and dangerous," AI applications offer a means to avoid putting human lives at risk or assigning humans to tasks that do not require the creativity of the human brain. AI systems also have the potential to reduce costs in logistics and sensing and to enhance communication and transparency in complex systems, if that is prioritized as a design value. In particular, as an information-communication technology, AI might benefit the peacekeeping agenda by more effectively communicating the capacities and motivations of military actors.

Workshop participants noted that AI-enabled systems and platforms have already made remarkable and important enhancements to national intelligence, surveillance, and reconnaissance capabilities. The ability of AI to support capturing, processing, storing, and analyzing visual and digital data has increased the quantity, quality, and accuracy of information available to decision makers. They can use this information to do everything from optimizing equipment maintenance to minimizing civilian harm. Additionally, these platforms allow for data capture in environments that are inaccessible to humans.

Participants shared broad agreement that the benefits of military applications of AI will require governments and the private sector to collaborate frequently and in depth. Specifically, participants advocated for the identification of practices and norms that ensure the safety of innovation in AI, especially in the testing and deployment phases. Examples include industry-level best practices in programming, industry and government use of test protocols, and government transparency and communication about new AI-based military capabilities.

Agreement also emerged over the need for better and more-comprehensive training among technologists, policymakers, and military personnel. Participants expressed clearly that managing the risks of AI will require technical specialists to have a better understanding of international relations and of the policymaking context. Effective policymaking and responsible use will also require government and military officials to have some knowledge of how AI systems work, their strengths, their possibilities, and their vulnerabilities. Practical recommendations for moving in this direction included the development of common terms for use in industry, government, and multilateral discourse, and including the private sector in weapons-review committees.

Potential Governance of Military Applications of AI

The primary challenge to multilateral governance of military AI is uncertainty—about the ways AI will be applied, about whether current international law adequately captures the problems that use of AI might generate, and about the proper venues through which to advance the development of governance approaches for military applications of AI. These characteristics of military AI are amplified by the technology's rapid rate of change and by the absence of standard and accepted definitions. Even fundamental concepts like autonomy are open to interpretation, making legislation and communication difficult.

There was skepticism among some, though not all, participants that current international law is sufficient to govern every possible aspect of the military applications of AI. Those concerned about the extent to which today's governance mechanisms are sufficient noted that there are specific characteristics of military applications of AI that may fit poorly into standing regimes—for example, international humanitarian law—or for which applying standing regimes may produce unintended consequences. This observation led to general agreement among participants that many governance approaches—including self-regulation, transparency and confidence-building measures, and intergovernmental approaches—ultimately would be required to mitigate the risks of military applications of AI. It should be noted that workshop participants included transnational nongovernmental organizations and transnational corporations—entities that increasingly have diplomatic roles.

The workshop concluded with general agreement that the UN system offers useful platforms within which to promote productive dialogue and through which to encourage the development of possible governance approaches between disparate stakeholders. All participants expressed the belief that beyond discussions on LAWS, broad understanding of and discourse about potential military applications of AI—its benefits, risks, and governance challenges—is nascent and, indeed, underdeveloped. Participants welcomed and encouraged more opportunities for stakeholders to educate each other, to communicate, and to innovate around the hard problems posed by military applications of AI.

Endnote

[1] *Transforming Our World: The 2030 Agenda for Sustainable Development*, Sustainable Development Goals Knowledge Platform, United Nations, accessed November 22, 2019, https://sustainabledevelopment.un.org/post2015/transformingourworld.

About the Author

Melanie Sisson is a Nonresident Fellow, Defense Strategy and Planning Program, with the Stimson Center.

This working paper was prepared for a workshop, organized by the Stanley Center for Peace and Security, UNODA, and the Stimson Center, on The Militarization of Artificial Intelligence.

Artificial Intelligence, Nuclear Weapons, and Strategic Stability

Jennifer Spindel | The University of New Hampshire

From a smart vacuum that can learn floor plans to "killer robots" that can revolutionize the battlefield, artificial intelligence has potential applications both banal and extraordinary. While applications in health care, agriculture, and business logistics can drive forward human development, military applications of artificial intelligence might make war more likely and/or increase its lethality. In both fact and science fiction, many of these new technologies are being developed by the private sector, introducing new governance challenges and stakeholders to conversations about the implications of new weapons development.

To begin addressing challenges related to the militarization of artificial intelligence, the Stanley Center for Peace and Security, in partnership with the United Nations Office for Disarmament Affairs and the Stimson Center, commissioned working papers from authors Paul Scharre, Vadim Kozyulin, and the China Arms Control and Disarmament Association (CACDA). This introductory paper provides background context to orient readers and highlights similarities and differences between those papers. It is organized around three primary sections: first, the difficulties in determining what artificial intelligence is or means; second, the ways artificial intelligence can affect the character of war in the broadest sense; and finally, the promises and pitfalls of applying artificial intelligence to nuclear weapons and systems.

C-3PO, Terminator, or Roomba: What Is Artificial Intelligence?

International discussions on artificial intelligence (AI) governance often revolve around the challenges of defining "artificial intelligence." AI is a diverse category that includes smart vacuums that learn floor plans and weapons that can acquire, identify, and decide to engage a target without human involvement. Defining what counts as AI, even in more-narrow military contexts, remains difficult. The working paper authors agree that artificial intelligence can mean many things and therefore has multiple applications to the military realm.

Paul Scharre notes that artificial intelligence is a general-purpose enabling technology, not unlike electricity. CACDA , in their paper, agree that AI will have wide applications to fields including agriculture, manufacturing, and health care, and he broadly defines artificial intelligence as the theories, methods, technologies, and application systems for stimulating, extending, and expanding human intelligence. While Vadim Kozyulin does not give a definition of AI, he explains that commercial companies, including Amazon, Microsoft, IBM, and Google, have created most artificial intelligence tools and then offered them to the military.

Because AI is not a single technology, the authors suggest various ways it could be applied to the military realm. Kozyulin, for example, points out that the Russian Ministry of Defense is interested in "combat robots." These robots are "multi-functional device[s] with anthropomorphic (humanlike) behavior, partially or fully performing functions of a person in executing a certain combat mission. [They include] a sensor system (sensors) for acquiring information, a control system, and actuation devices." CACDA and Scharre suggest less overtly militarized applications of AI, including intelligence, surveillance, and reconnaissance (ISR) operations, or actually analyzing and interpreting sensor data, or geospatial imagery analysis. Whether it is used for combat robots or analyzing data, artificial intelligence has the potential to decrease human involvement in war. As the next section discusses, this means AI could fundamentally change the character of war.

The Evolving Character of War

Though conflict carried out entirely, or even primarily, by combat robots is an unlikely scenario, the authors agree that artificial intelligence will affect war in at least two ways. First, artificial intelligence will affect military organizations and combat philosophy by changing the distribution of human and machine resources needed to engage in war and war-adjacent operations. Second, artificial intelligence will affect the speed of operations, which will, paradoxically, both increase and decrease the time for decision making. However, the authors also articulate concerns about importing AI-enabled technology into the military realm, particularly in terms of training AI and AI flexibility.

AI is likely, the authors believe, to affect military organizations and combat philosophy by freeing humans to focus on the things that really matter. Applied to ISR operations, AI could allow human analysts to focus on the data points or images that are potentially most meaningful, rather than spend the majority of their time sifting through thousands of status quo images. Delegating the "drudgery" tasks to an AI system holds much promise, the authors argue.

AI will also affect military organizations and combat philosophy through increased automation. Scharre explains that autonomous vehicles will become faster, stealthier, smaller, and more numerous, and will persist longer on the battlefield. This means the presence of humans on the battlefield could decrease, allowing them to focus on bigger strategic issues rather than fighting at the tactical level. CACDA argues that an increasingly automated and mechanized mode of fighting will make it possible to more accurately and reliably predict conflict outcomes. Combat philosophy could fundamentally change because nations would not fight unless they knew they would (or had a strong chance to) win. Ultimately, these changes to military organization and combat philosophy could imply stability and an overall decrease in war.

The second way AI will affect the character of war concerns changes to speed. If AI systems can act and react more quickly than humans, the reduced time for decision making increases the likelihood of accidental or misperceived escalation in a conflict. As all three authors point out, there are no referees to call time out in war. Militaries will need to balance their desire for a speedy response with the presence of circuit breakers to limit the potential consequences of actions. Kozyulin articulates concerns about "hyperwar" or "battlefield singularity," where the speed of AI-enabled actions outpaces human capacity for decision making, leading to less human control over war. This is of particular concern if, as Scharre explains, actions are predelegated to an AI system. During the Cuban Missile Crisis, US leaders changed their minds about whether to attack the Soviet Union. An AI system would not have the same ability to mull over its decisions and reverse course, as humans often do.

On the other hand, if AI systems can identify and classify things—for example, objects, threats, or individuals—more quickly than humans, there could be more time for humans to make decisions, which would decrease the risks of accidental or misperceived escalation. Whether speed is a net positive or a net detriment to wartime decision making will depend on how AI systems are integrated into military systems and what range of actions are predelegated.

However, the papers are careful to point out that artificial intelligence is not a panacea; there are many risks involved with applying AI to the military domain. Two key challenges are in training an AI system to operate in wartime environments and in programming morals or ethics into an AI system.

While humans can train on mock battlefields, the same is not true for AI. AI systems learn by doing, but mock battlefields don't come close enough to simulating real operational conditions. An AI system that drives autonomous cars, for example, is trained by driving millions of miles on public roads before it is allowed to drive on its own. There is no analogous situation for military autonomous systems: systems cannot be tested under real operational conditions until wartime. A machine that performs well in a training environment might fail during war, with consequences ranging from the minorly inconvenient to the catastrophic. In 2003, a US Patriot air defense system operating in automated mode shot down a British Royal Air Force plane, killing the pilot and the navigator. A few days later, it shot down a US Air Force plane, killing the pilot. Automated mode allows the Patriot to decide to launch missiles without human interaction.[1]

Training an AI algorithm is extremely important if it is to accurately and quickly analyze input data. Even before it is trained under mock wartime conditions, an AI system needs to learn, using prelabeled objects with positive and negative examples. To identify an object as a missile, for example, an algorithm needs to be able to distinguish "missile" and "not-missile" as distinct object categories. There are not many pictures of mobile missile launchers available publicly, which raises concerns about the risk of false negatives.[2] The stakes for getting this right are high: if an AI system identifies a target as a military one, but the target is actually a school bus or a hospital, tragedy will follow. This type of mistake has happened when humans who have extensive battle training are involved, and that suggests that the issue of training an AI system is one of the crucial obstacles to incorporating AI into the military realm.

Finally, the authors raise questions about morals, ethics, and legal thresholds for using AI systems in war. Kozyulin notes that there is a general lack of discussion about these issues and is concerned that the Russian defense industry's focus on closing the technological gap means AI is treated as a technical task, with no room for deeper thinking about the moral or philosophical dimensions. CACDA similarly notes that the ethical issues require proficiency in law and philosophy, which aren't usually taught or required in many of the more technical fields. Technical and moral/ethical issues become more acute when we consider the applications of AI to nuclear weapons and strategy.

When AI Meets Nuclear Weapons

Concerns about how artificial intelligence can affect the character of war are amplified in the nuclear realm. The authors agree that allowing artificial intelligence systems to make launch decisions is probably not a prudent idea but believe there is promise in using AI to improve nuclear early warning systems.

If there is a lack of wartime data to train AI systems in conventional battle, there is even less data for training AI to make nuclear-related decisions. Although there is evidence from nearly seventy years of aircraft, submarine, and missile tests of nuclear weapons, an effective and reliable AI system would need significantly more data—real and simulated—for training purposes. The problem of training an AI system in the nuclear realm can be illustrated by looking at near-launch decisions made by humans.

Scharre provides a detailed overview of the familiar Petrov case. As many readers will know, in September 1983, the Soviet missile alert system detected five incoming US intercontinental ballistic missiles (ICBMs). Thankfully, Lt. Col. Stanislav Petrov, the Soviet officer on duty, thought that five was an unusually small number of missiles with which to initiate nuclear apocalypse and reported that the Soviet system was malfunctioning.[3] How can an AI system learn to differentiate five real ICBMs from sunlight reflecting off of the clouds? While there has not yet been accidental detonation of a nuclear weapon, the track record on nuclear safety—including accidents, almost-launches, and missing nuclear warheads—does not inspire confidence.[4] Consider again the case of the US Patriot missiles that shot down friendly planes in 2003. If that type of launch authority were given to nuclear weapons, an accident could quickly prove catastrophic, with worldwide consequences.

On the other hand, it's possible that an AI system would have flagged the five ICBMs as anomalous, just as Petrov did, and worth further human investigation. That possibility is why the authors agree there is promise in applying artificial intelligence to nuclear early warning systems. Scharre says that rather than causing false alarms, an early warning system equipped with artificial intelligence capabilities could give accurate warnings of surprise attacks.

Beyond surprise attacks, AI could create time efficiencies during crises by speeding up processes and augmenting decision making. If AI can improve the speed and quality of information processing in advanced nuclear warning systems, decision makers could have more time to react, which could slow down a crisis situation. Time is a key commodity in a nuclear crisis, since a nuclear-armed missile could reach its target in as little as eight minutes.[5] AI-enabled early warning systems could be crucial in opening an otherwise tightening window of decision, relieving some of the pressure to act immediately.

A similar potential benefit of bringing AI into the nuclear realm would be using it to interpret and monitor sensor data. Similar to ISR operations, this would use AI systems to detect anomalies or change—in reactors, inventories, nuclear materials movement, etc.—and direct human focus for further investigation.

However, AI presents three main risks to deterrence and nuclear stability. First, AI has opened up conversations about counterforce targeting, which increases the likelihood of misperception and miscalculation. Kozyulin suggests that AI could enable precision strikes to destroy key command, control, and communication assets. He also suggests AI could be used to make the oceans more transparent, which raises questions about the invulnerability of nuclear submarines. If nuclear assets in previously hard-to-detect places become trackable, then countries will face an unstable situation not seen since the early days of the Cold War. Crucially, it doesn't matter whether AI actually enables these capabilities; the mere perception that AI puts counterforce targeting in reach is destabilizing.

Second, AI can undermine deterrence by inducing states to deliberately use their nuclear weapons. If states believe that counterforce targeting is possible, or that other states could use AI systems to interfere with their command and control, they might feel pressured to use their weapons now.[6] The decision-making logic unfolds as follows: nations want to be assured that their weapons will always fire when so ordered and will never fire unless the launch is intentional. The possibility that a submarine could be destroyed, or its systems hacked, complicates this always-never calculation. Fearing that they might eventually lose the ability to use their nuclear weapons, states might decide to use them now, rather than risk future obsolescence.

Though CACDA suggests that AI could revolutionize war by enabling more precise and certain calculations about costs and benefits, precision and certainty are problematic concepts in the nuclear realm. Uncertainty is a feature, not a bug, where nuclear weapons are concerned.

Finally, as states seek to include AI in ever higher levels of nuclear decision making, the risks for accidents also increase. Kozyulin warns that using AI in nuclear early warning systems could lead states to militarize at ever higher technological levels. This would increase tensions and lead to arms races. Scharre similarly cautions against a "race to the bottom on safety." As states feel pressure to keep up and innovate, they may deploy AI systems and/or new weapons that haven't been fully tested, which increases the risks of accidents.

Conclusion: Artificial Intelligence and Strategic Stability

The papers demonstrate the unknowns about how AI will affect war, nuclear war, and strategic stability. Many of them take the format of "if X, then Y," since the future of AI is unknown. The papers suggest many reasons to be concerned, and some reasons to be optimistic, about the development of AI and its application to the military realm.

Overall, there is concern about strategic stability. While AI might enable increased accuracy and interpretation of sensor and imagery data, the authors share concerns about predelegation and the inflexibility of AI systems, which increase the risks of an accident or miscalculated use. If a situation is strategically stable where war only occurs if one side truly seeks it—and not because of accidents or misperception—the jury is out on the effects of AI on strategic stability.[7]

One area of concern is asymmetric strategies. Kozyulin notes the large investment needed to keep pace with AI technical developments. If a direct arms race isn't possible because some nations can't keep pace with technologies or their costs, then he suggests they are likely to turn to asymmetric responses. For example, if one country bolsters its counterforce targeting by integrating AI systems with its nuclear command and control, an adversary might counter this by raising its alert rates and predelegating launch authority. While predelegation might temporarily bolster deterrence, it ultimately increases the risks of accidental or misperceived nuclear use.[8]

For states that do have the budget to invest in AI innovation, all three authors are concerned about continued competition and arms race dynamics. While Scharre expressed concerns about militaries taking safety short cuts, Kozyulin and CACDA envision future competition between heavily automated attack and defense systems. They both name hypersonic missiles—which Russia is currently developing—as a concern. CACDA also believes that certain areas of warfare, such as cyberspace and electromagnetic, could become completely outsourced to nonhuman weapons. There is reason to believe that this competitive dynamic is already in play. In September 2017, Russian President Vladimir Putin said, "Artificial intelligence is not only the future of Russia, it is the future of all mankind. There are enormous opportunities and threats that are difficult to predict today. The one who becomes a leader in this sphere will be the ruler of the world."[9] This type of competitive dynamic could lead to greater instability in a multipolar world. While Cold War competition saw the United States and Soviet Union as the two centers of gravity, today more states are developing and capable of developing AI systems, which means that more rivalries and competitive dynamics are in play. AI development is therefore likely to be critical for the macro balance of power.[10]

In the face of such unknowns, the papers offer suggestions for developing a more cooperative future. Though unsure about the feasibility of arms control measures, Scharre and Kozyulin suggest more research into governance and treaty protocols that could be used to regulate AI. CACDA is more skeptical and notes that there have been more that 40 proposals for AI ethics guidelines already. He suggests more-pragmatic principles for governing autonomous weapons and AI, including the legal threshold for the use of force.

CACDA also reminds us of the promises of AI in other domains and wants to ensure that the potential risks of militarized AI do not prevent the realization of gains in other areas. CACDA notes that places where AI is most needed, like health care, often do not get direct funding priorities. Nor does research on AI and ethics. However, The paper from CACDA asks us not to lost sight of the ways AI could improve global productivity and economic development in sectors as diverse as health care, agriculture, and infrastructure development. Cooperation and collaboration in AI development in these areas could lead to global scientific progress and innovation.[11]

The papers discuss a wide range of political and technical developments concerning artificial intelligence. Connecting all of them is the often-unstated importance of human perceptions. One of the key issues for understanding AI and deterrence is figuring out how to convince others that AI will not be used for counterforce targeting. That question is ultimately one of perceptions and psychology, rather than technical developments. Like previous instances of technological innovation, the effects of AI and political and military development will depend on how people, organizations, and societies decide to adopt and use technologies.[12] The raw characteristics of AI offer a number of plausible futures, but Scharre, Kozyulin, and CACDA demonstrate the importance of human decisions about how to use this new technology.

Endnotes

[1] Connor McLemore and Charles Clark, “The Devil You Know: Trust in Military Applications of Artificial Intelligence,” War on the Rocks, September 23, 2019, https://warontherocks.com/2019/09/the-devil-you-know-trust-in-military-applications-of-artificial-intelligence/?fbclid=IwAR15Wczbr9uLN0nFhYkIW1QFyQrkA7F9Ql0UqIrhDiAO5LPL0Yku81Jn9ZQ.

[2] Rafael Loss and Joseph Johnson, “Will Artificial Intelligence Imperil Nuclear Deterrence?” *War* on the Rocks, September 19, 2019, https://warontherocks.com/2019/09/will-artificial-intelligence-imperil-nuclear-deterrence/?fbclid=IwAR1OUa90LH0jKAuISTpRvycpF7-FIHfYuwZRVm-07WhKDEdzH1L4J7YfjKQ.

[3] Paul Scharre, “Autonomous Weapons and Operational Risk,” Center for a New American Security, February 2016, https://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515.

[4] Scharre, “Autonomous Weapons,” 51; “Accidents, Errors, and Explosions,” *Outrider Post*, https://outrider.org/nuclear-weapons/timelines/accidents-errors-and-explosions/.

[5] Jaganth Sankaran, “A Different Use for Artificial Intelligence in Nuclear Weapons Command and Control,” War on the Rocks, April 25, 2019, https://warontherocks.com/2019/04/a-different-use-for-artificial-intelligence-in-nuclear-weapons-command-and-control/; Ariel Conn, “Podcast: AI and Nuclear Weapons—Trust, Accidents, and New Risks with Paul Scharre and Mike Horowitz,” Future of Life Institute, September 28, 2018, https://futureoflife.org/2018/09/27/podcast-ai-and-nuclear-weapons-trust-accidents-and-new-risks-with-paul-scharre-and-mike-horowitz/?cn-reloaded=1&cn-reloaded=1.

[6] Jonathan Clifford, “AI Will Change War, but Not in the Way You Think,” War on the Rocks, September 2, 2019, https://warontherocks.com/2019/09/ai-will-change-war-but-not-in-the-way-you-think/.

[7] Elbridge Colby, *Strategic Stability: Contending Interpretations* (Carlisle, PA: US Army War College Press, 2013), 57.

[8] Mansoor Ahmed, “Pakistan’s tactical Nuclear Weapons and Their Impact on Stability,” Carnegie Endowment for International Peace, June 30, 2016, https://carnegieendowment.org/2016/06/30/pakistan-s-tactical-nuclear-weapons-and-their-impact-on-stability-pub-63911; Vipin Narang, “Posturing for Peace? Pakistan’s Nuclear Postures and South Asian Stability,” *International Security* 34, no. 3 (Winter 2009/10): 38-78.

[9] Vadim Kozyulin, “Militarization of AI,” The Militarization of Artificial Intelligence, June 2020, 25.

[10] Michael Horowitz, Elsa B. Kania, Gregory C. Allen, and Paul Scharre, “Strategic Competition in an Era of Artificial Intelligence,” Center for aNew American Security, 25 July 2018, https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence.

[11] Kania, “The pursuit of AI is more than an arms race.”

[12] Michael Horowitz, “Artificial Intelligence, International Competition, and the Balance of Power,” *Texas National Security Review* 1, no. 3, May 2018, https://doi.org/10.15781/T2639KP49.

About the Author

Jennifer Spindel is an Assistant Professor of Political Science, University of New Hampshire..

This working paper was prepared for a workshop, organized by the Stanley Center for Peace and Security, UNODA, and the Stimson Center, on The Militarization of Artificial Intelligence.

Discussion Paper | A US Perspective

Military Applications of Artificial Intelligence:

Potential Risks to International Peace and Security

Paul Scharre | Center for a New American Security

July 2019

Recent years have seen an explosion in the possibilities enabled by artificial intelligence (AI), driven by advances in data, computer processing power, and machine learning.[1] AI is disrupting a range of industries and has similar transformative potential for international relations and global security. At least two dozen countries have released national plans to capitalize on AI, and many states are seeking to incorporate AI to improve their national defense.[2] This paper aims to improve understanding of how militaries might employ AI, where those uses might introduce risks to international peace and security, and how states might mitigate these risks.[3]

Artificial intelligence is not a discrete technology like a fighter jet or locomotive, but rather is a general-purpose enabling technology, like electricity, computers, or the internal combustion engine. As such, AI will have many uses. In total, these uses could lead to economic growth and disruption on the scale of another industrial revolution. This AI-driven cognitive revolution will increase productivity, reduce automobile accidents, improve health outcomes, and improve efficiency and effectiveness in a range of industries. Many, but not all, of the recent advances in AI come from the field of machine learning, in which machines learn from data, rather than follow explicit rules programmed by people.[4] AI continues to advance as a field of study,[5] but even if all progress were to stop today (which is unlikely),[6] there would still be many gains across society by applying current AI methods to existing problems.

The net effect of AI across society is likely to be very beneficial, but both malign and responsible actors will use AI in security applications as well. Better understanding these uses, and how to counter them when necessary, is essential to ensuring that the net effect of AI on society is maximally beneficial. State and nonstate actors have already caused harm through the deliberate malicious use of AI technology. As AI technology moves rapidly from research labs to the real world, policy makers, scholars, and engineers must better understand the potential risks from AI in order to mitigate against harm.[7]

War + AI

As a general-purpose enabling technology, AI has many potential applications to national defense. Military use of AI is likely to be as widespread as military use of computers or electricity. In the business world, technology writer Kevin Kelly has said, “There is almost nothing we can think of that cannot be made new, different, or interesting by infusing it with” greater intelligence. To imagine business applications, “Take X and add AI.”[8] The same is true for military AI applications. AI is likely to affect strategy, operations, logistics, personnel, training, and every other facet of the military. There is nothing intrinsically concerning about the militarization of artificial intelligence, any more than the militarization of computers or electricity is concerning. However, some specific military applications of AI could be harmful, such as lethal autonomous weapons or the application of AI to nuclear operations. Additionally, the net effect of the “intelligentization” or “cognitization” of military operations could alter warfare in profound ways.[9]

The first and second Industrial Revolutions dramatically changed warfare, increasing the scope and scale of destruction that could be inflicted with industrial-age weapons. Policy makers at the time were unprepared for these changes, and the result was two global wars with tens of millions of lives lost. This increased scale of destruction was not due to one or two specific uses of industrial technology in war but rather the net effect of industrialization. The Industrial Revolutions enabled the mass mobilization of entire societies for “total war,” as nations turned the increased productivity and efficiency made possible by industrial technology to violent ends. Steel and the internal combustion engine made it possible to build war machines like the tank, submarine, and airplane and to take warfare to new domains under the sea and in the air. Mechanization enabled an expansion of destructive capacity through weapons like the machine gun, leading to the deadly trench warfare of World War I. And radio communications enabled coordinated long-distance operations, making possible lightning advances like the *blitzkrieg* of World War II.

As warfare transitioned to the Atomic Age, the extreme destructive potential of nuclear weapons was made clear in the aftermath of the bombings of Hiroshima and Nagasaki. Policy makers understood the stakes of nuclear-era warfare and the existential risk it posed—and still poses—to humanity. Yet the effect of AI on warfare is more likely to be similar to that of the Industrial Revolution, with myriad changes brought about by the widespread application of general-purpose technologies, rather than a single discrete technology like nuclear weapons.

Industrialization increased the physical scope and scale of warfare, allowing militaries to field larger, more-destructive militaries that could move farther and faster, delivering greater firepower, and in a wider array of domains. Artificial intelligence is bringing about a cognitive revolution, and the challenge is to anticipate the broad features of how this cognitive revolution may transform warfare.

Features of Artificial Intelligence

Value of AI Systems

The field of artificial intelligence comprises many methods, but the goal is to create machines that can accomplish useful cognitive tasks.[10] Today’s AI systems are narrow, meaning they are only capable of performing the specific tasks for which they have been programmed or trained. AI systems today lack the broad, flexible general intelligence that humans have that allows them to accomplish a range of tasks. While AI methods are general purpose and can be applied to solve a wide range of problems, AI systems are not able to flexibly adapt to new tasks or environments on their own. Nevertheless, there are many tasks for which AI systems can be programmed or trained to perform useful functions, including in many cases at human or even superhuman levels of performance. AI systems do not always need to reach superhuman performance to be valuable, however. In some cases, their value may derive from being cheaper, faster, or easier to use at scale relative to people.

Some of the things AI systems can do include classifying data, detecting anomalies, predicting future behavior, and optimizing tasks. Real-world examples include AI systems that:

- **Classify** data, from song genres to medical imagery.
- **Detect** anomalous behavior, such as fraudulent financial transactions or computer malware.
- **Predict** future behavior based on past data, such as recommendation algorithms for media content or better weather predictions.
- **Optimize** performance of complex systems, allowing for greater efficiency in operations.

In military settings, provided there was sufficient data and the task was appropriately bounded, in principle, AI systems may be able to perform similar tasks. These could include classifying military objects, detecting anomalous behavior, predicting future adversary behavior, and optimizing the performance of military systems.

Autonomy

Artificial intelligence can also enable autonomous systems that have greater freedom to perform tasks on their own, with less human oversight. Autonomy can allow for superhuman precision, reliability, speed, or endurance. Autonomy can also enable greater scale of operations, with fewer humans needed for large-scale operations. Autonomy can allow one person to control many systems. When embedded into physical systems, autonomy can allow vehicles with forms that might be impossible if humans were onboard, or operation in remote or dangerous locations. Autonomy enables robot snakes that can slither through pipes, underwater gliders that can stay at sea for years at a time, swarms of small expendable drones, and robots that can help clean up nuclear disasters.

Limitations of AI Systems Today

Artificial intelligence has many advantages, but it also has many limitations.[11] Today’s AI systems fall short of human intelligence in many ways and are a far cry from the Cylons, Terminators, and C-3POs of science fiction.

One of the challenges of AI systems is that the narrowness of their intelligence means that while they may perform very well in some settings, in other situations their performance can drop off dramatically. A self-driving car that is far safer than a human driver in one situation may suddenly and inexplicably drive into a concrete barrier, parked car, or semitrailer.[12] A classification algorithm that performs accurately in one situation may do poorly in another. The first version of AlphaGo, which reached superhuman performance in 2016, reportedly could not play well if the size of the game board was changed from the 19-by-19-inch board on which it was trained.[13] The narrow nature of AI systems makes their intelligence brittle—susceptible to sudden and extreme failure when pushed outside the bounds of their intended use.

Failures can manifest in a variety of ways. In some cases, the system’s performance may simply degrade. For example, a facial-recognition algorithm trained on people of one skin tone may perform less accurately on people of a different skin tone.[14] In other circumstances, a failure may manifest more dramatically, such as a self-driving car that suddenly attempts to drive through an obstacle. Some failures may be obvious, while others may be more subtle and escape immediate detection but nevertheless result in suboptimal outcomes. For example, a resume-sorting AI system may have a subtle bias against certain classes of individuals.[15] Because of the opaque nature of machine learning systems, it may be difficult to understand why a system has failed, even after the fact.

One complicating factor for increasingly sophisticated AI systems is that their complexity makes them less transparent to human users. This means that it can be more difficult to discern when they might fail and under what conditions. For very complex systems operating in real-world environments, there is a seemingly infinite number of possible interactions between the system's programming and its environment.[16] It is impossible to predict them all. Computer simulations can help expand the scenarios a system is evaluated against, but testers are still limited by what they can imagine, and even the best simulations will never perfectly replicate the real world. Self-driving-car companies are simulating millions of driving miles every day with computers, and still there will be situations in the real world they could not have anticipated, some of which may cause accidents.[17]

AI systems are also vulnerable to a range of cognitive attacks that are analogous to cyberattacks but work at the cognitive level, exploiting vulnerabilities in how the AI system "thinks." Examples include poisoning the data used to train an AI system or adversarial attacks that spoof AI systems with tailored data inputs, causing them to generate incorrect outputs.[18]

All of these limitations are incredibly relevant in military environments, which are chaotic, unpredictable, and adversarial. Militaries will use AI systems, and those AI systems will break. They will suffer accidents, and they will be manipulated intentionally by adversaries. Any assessment of the role of AI in warfare must take into account the extreme brittleness of AI systems and how that will affect their performance on the battlefield.

War in the Cognitive Age

Artificial intelligence will introduce a new element to warfare: supplementing and augmenting human cognition. Machines, both physical and digital, will be able to carry out tasks on their own, at least within narrow constraints. Because today's AI systems are narrow, for the foreseeable future human intelligence remains the most advanced cognitive processing system on the planet. No AI system, or even suite of systems, can compare with the flexibility, robustness, and generality of human intelligence. This weakness of machine intelligence and strength of human intelligence is particularly important in warfare, where unpredictability and chaos are central elements. Warfare in the cognitive age will be partly a product of AI but also of human intelligence, which will remain a major feature of warfare for the foreseeable future.

Even though humans will remain involved, the introduction of artificial intelligence is likely to dramatically change warfare. AI will enable the fielding of autonomous vehicles that are smaller, stealthier, faster, more numerous, able to persist longer on the battlefield, and take greater risks.[19] Swarming systems will be valuable for a range of applications, including reconnaissance, logistics, resupply, medical evacuation, offense, and defense.

The most profound applications of AI are likely to be in information processing and command and control. Just as industrialization changed the physical aspects of warfare, artificial intelligence will principally change the cognitive aspects of warfare. Militaries augmented with AI will be able to operate faster and with more-numerous systems, and conduct more-complex and distributed operations.

While much of the attention on military AI applications has focused on robotics, it is worth noting that in computer games, such as Dota 2, computers have achieved superhuman performance while playing with the same units as human competitors.[20] Computers' advantages have come in better and faster information processing, and command and control. Whereas humans can only pay attention to a limited number of things, an AI system can simultaneously absorb and process all incoming information at once. Machines can then process this information faster than humans and coordinate the simultaneous rapid responses of military units. These advantages will make AI systems valuable for militaries in improving battlefield awareness, command and control, and speed, precision, and coordination in action. Because of machines' limitations in responding to novel situations, however, humans will still be needed in real-world combat environments, which are more complex and unrestricted than computer games. The most effective militaries are likely to be those that optimally combine AI with human cognition in so-called centaur approaches, named after the mythical half-human, half-horse creature.

Potential Risks from Military AI Applications

The introduction of AI could alter warfare in ways both positive and negative. It can be tempting to envision AI technologies as principally enabling offensive operations, but they will be valuable for defensive operations as well. Because AI is a general-purpose technology, how it shifts the offense-defense balance in different areas may depend on the specific application of AI, and may evolve over time.

Some general characteristics of AI and attendant risks are outlined below, but it is worth noting that these risks are only possibilities. Technology is not destiny, and states have choices about how to use AI technology. How these risks manifest will depend on what choices states make. A concerted effort to avoid these risks may be successful.

Accident Risk

In principle, automation has the potential to increase precision in warfare and control over military forces, reducing civilian casualties and the potential for accidents that could lead to unintended escalation. Automation has improved safety in commercial airline autopilots and, over time, will do so for self-driving cars. However, the challenge in achieving safe and robust self-driving cars in all weather and driving conditions points to the limitations of AI today. War is far more complex and adversarial than driving or commercial flying.

An additional problem militaries face is a lack of available data on the wartime environment. To build self-driving cars that are robust to a range of driving conditions, the autonomous car company Waymo has driven over 10 million miles on public roads. Additionally, it is computer simulating 10 million driving miles every day.[21] This allows Waymo to test its cars under a variety of conditions. The problem for militaries is that they have little to no ground-truth data about wartime conditions on which to evaluate their systems. Militaries can test their AI systems in training environments, either in the real world or in digital simulations, but they cannot test their actual performance under real operational conditions until wartime. Wars are a rare occurrence, fortunately. This poses a problem for testing autonomous systems, however. Militaries can do their best to mimic real operational conditions as closely as possible in peacetime, but they can never fully recreate the chaos and violence of war. Humans are adaptable and are expected to innovate in wartime, using their training as a foundation. But machine intelligence is not as flexible and adaptable as human intelligence. There is a risk that military AI systems will perform well in training environments but fail in wartime because the environment or operational context is different, perhaps even only slightly different. Failures could result in accidents or simply cause military systems to be ineffective.

Accidents with military systems could cause grave damage. They could kill civilians or cause unintended escalation in a conflict. Even if humans regained control, an incident that killed adversary troops could escalate tensions and inflame public sentiment such that it was difficult for national leaders to back down from a crisis. Accidents, along with vulnerabilities to hacking, could undermine crisis stability and complicate escalation management among nations.

Autonomy and Predelegated Authority

Even if AI systems perform flawlessly, one challenge nations could face is the inability to predict themselves what actions they might want to take in a crisis. When deploying autonomous systems, humans are predelegating authority for certain actions to a machine. The problem is that in an actual crisis situation, leaders may decide that they want to take a different approach. During the Cuban Missile Crisis, US leaders decided that if the Soviets shot down a US reconnaissance plane over Cuba, they would attack. After the plane was shot down, they changed their minds. Projection bias is a cognitive tendency where humans fail to accurately predict their own preferences in future situations. The risk is that autonomous systems perform as programmed, but not in ways that human leaders desire, raising the risk of escalation in crises or conflicts.

Prediction and Overtrust in Automation

Maintaining humans in the loop and restricting AI systems to only giving advice is no panacea for these risks. Humans frequently overtrust in machines, a phenomenon known as automation bias.[22] Humans were in the loop for two fratricide incidents with the highly automated US Patriot air and missile defense system in 2003 yet failed to stop the accidents.[23] In one notable psychological experiment, participants followed a robot the wrong way through a smoke-filled building that was simulating a fire emergency, even after being told the robot was broken.[24]

Overtrusting in machines could lead to accidents and miscalculation, even before a war begins. In the 1980s, the Soviet Union conducted Operation RYaN to warn of a surprise US nuclear attack. The intelligence program tracked data on various potential indicators of an attack, such as the level of blood in blood banks, the location of nuclear weapons and key decisionmakers, and the activities of national leaders.[25] If AI systems could actually give accurate early warning of a surprise attack, this could be stabilizing. Knowing that there was no possibility of successfully carrying out a surprise attack, nations might refrain from attempting one. Yet prediction algorithms are only as good as the data on which they are trained. For rare events like a surprise attack, there simply isn't enough data available to know what is actually indicative of an attack. Flawed data will lead to flawed analysis. Yet the black-box nature of AI, in which its internal reasoning is opaque to human users, can mask these problems. Without sufficient transparency to understand how the algorithm functions, human users may not be able to see that its analysis has gone awry.

Nuclear Stability Risks

All of these risks are especially consequential in the case of nuclear weapons, where accidents, predelegated authority, or overtrust in automation could have grave consequences. False alarms in nuclear early warning systems, for example, could lead to disaster. There have been numerous nuclear false alarms and safety lapses with nuclear weapons throughout the Cold War and afterward.[26] In one particularly notable incident in 1983, a Soviet early warning satellite system called Oko falsely detected a launch of five US intercontinental ballistic missiles against the Soviet Union. In fact, the satellites were sensing the reflection of sunlight off of cloud tops, but the automated system told human operators "missile launch." Soviet Lieutenant Colonel Stanislav Petrov judged the system was malfunctioning, but in future false alarms, the complexity and opacity of AI systems could lead human operators to overtrust those systems.[27] The use of AI or automation in other aspects of nuclear operations could pose risks as well. For example, nuclear-armed uninhabited aircraft (drones) could suffer accidents, leading states to lose control of the nuclear payload or accidentally signaling escalation to an adversary.

Competitive Dynamics and Security Dilemmas

Competition exacerbates many of these risks. Despite media headlines warning of an AI arms race, the current situation among states does not resemble previous arms races, in which countries spent escalating sums of money on battleships or nuclear weapons without gaining any clear military advantage. AI innovation today is largely driven by the commercial sector, and militaries seek to import AI technology to defense applications. Competitive dynamics could still lead to security dilemmas, in which states individually take actions to increase their own security, but with the net effect of decreasing security for all. The two greatest risks in a race to use AI are in speed and safety.

Speed

One of the great dangers of automation is an arms race in speed, in which countries push humans further and further out of the loop in a bid to act faster than competitors. The consequences of this dynamic can be seen in stock trading, which is highly automated today. Algorithms execute trades at speeds measured in microseconds (1 microsecond equals 0.000001 seconds).[28] In a single eyeblink, 100,000 microseconds pass by. Yet when algorithms get it wrong, they can wreak havoc at machine speed. In the May 2010 "flash crash," a combination of brittle algorithms, high-frequency trading, market instability, and humans taking advantage of predictable bot behavior all combined to create a perfect storm in which the US stock market lost nearly 10 percent of its value in minutes.[29] Two years later, the high-frequency trading firm Knight Capital Group suffered an accident with a runaway algorithm, which began making erroneous trades at machine speed, moving $2.6 million a second. Within 45 minutes, it had lost $460 million, more than the company's entire assets.[30]

Financial regulators have dealt with the problem of flash crashes not by preventing them from occurring but by installing circuit breakers that take stocks offline if prices move too quickly and mitigate the consequences of an event.[31] Miniflash crashes continue to occur, and in a single day in 2015, over 1,200 circuit breakers were tripped across multiple financial markets around the globe.[32]

An escalatory incident between competitive military AI systems could have serious consequences. The challenge nations face is that there are no referees to call time out in war. If nations are to prevent such an incident, they will need to build in their own circuit breakers to limit the potential consequences of automation. These risks are particularly acute in cyberspace, where cybersystems could have global effects in seconds. A flash war would benefit no one.

Even once a war begins, an AI-accelerated operational tempo could lead to less human control over battlefield operations. Some Chinese scholars have hypothesized about a "battlefield singularity" in which the pace of action eclipses human decision making, and some US scholars have used the term "hyperwar" to refer to a similar situation.[33] The problem is that greater speed on one side necessitates greater speed on the other, with a net outcome that is more harmful for all. Moving to a new domain of warfare with less human control would be dangerous and risk large-scale accidents or escalation, even within a conflict. All militaries have an incentive to keep war more effectively under human control.

Race to the Bottom on Safety

Speed is not only a concern on the battlefield but also in peacetime development and deployment of military systems. Testing and evaluation are vitally important for improving the safety of complex autonomous systems. Greater testing in real-world and simulated environments can help identify flaws in a system ahead of time and reduce the risk of accidents. While no amount of testing can render a system 100 percent accident proof, more-extensive testing can help reduce the risk of accidents.

Unfortunately, a desire to beat a competitor to fielding a new system could cause actors to cut corners on safety, deploying autonomous systems before they are ready. This speed-to-market dynamic has been implicated as a possible contributing factor to accidents in the commercial airline autopilot industry and self-driving cars. If such a dynamic were to befall militaries, the result would be a world of unreliable military AI systems, which would make all nations less safe.[34]

Mitigating Potential Risks

Nations build militaries precisely because they don't trust others and want to provide for their own defense. In spite of this, states have come together on many occasions to limit the proliferation, development, production, stockpiling, or use of various military technologies that were seen as excessively harmful, inhumane, or destabilizing. Arms control is one option for mitigating risks from AI, but there are other unilateral measures states can take.

Technology Controls

Military technologies can be controlled or restricted at a number of stages along their development cycle. Nonproliferation regimes aim to limit access to the underlying technology behind certain weapons. The Nuclear Non-Proliferation Treaty, for example, aims to prevent the spread of nuclear weapons, promote cooperation on peaceful uses of nuclear energy, and further the goal of nuclear disarmament. Some weapons bans, like those on land mines and cluster munitions, allow access to the technology but prohibit developing, producing, or stockpiling the weapon. Other bans only apply to use, sometimes prohibiting use entirely or proscribing only certain kinds of uses of a weapon. Finally, arms-limitation treaties permit use but limit the quantities of certain weapons states can have in peacetime.[35]

AI is not like nuclear technology; it is more like computers, which are diffuse and driven by the commercial sector.[36] AI research papers are openly published online, and trained AI models can often be downloaded for free from online resources. Many actors will have access to AI, and preventing the underlying availability of AI is not likely to be feasible, at least given the shape of AI technology today. However, the specific uses of AI are more important, and states have choices about how the technology is used. Bans on land mines and cluster munitions don't prohibit access to the technology, but they do prohibit producing, stockpiling, or using those weapons. Arms control over AI as a whole would likely be infeasible, like attempting arms control for industrialization.

However, the Industrial Revolution saw a raft of treaties on various applications of industrial technology to war, treaties that had a mixed track record of success in the late 19th and early 20th centuries. Similarly, it is possible to conceive that arms control on some applications of AI could be successful. Achieving trust among all parties would be challenging, since AI systems are software and not observable in the same way naval ships or nuclear missiles are, which permits states to verify that others are complying with the treaty. However, there may be ways to achieve sufficient verification and compliance through other means or on some aspects of AI.

Transparency and confidence-building measures could also help reduce the risk of accidents by reducing the potential for miscalculation or misunderstanding among states.[37]

Building Safe and Secure AI Systems

Ultimately, the most powerful tool states have at their disposal for mitigating the risk of military AI systems comes from building safe and secure AI systems themselves. Militaries have an incentive to keep their systems under effective operational control. AI systems that slip out of human control could not only cause an accident, possibly harming third parties, but are also not very useful to the military that deploys them. Military systems that may not work or could be hacked by the enemy are not very useful or valuable. Conducting better tests and evaluation and maintaining humans in overall operational control of the system through a human-machine centaur command-and-control model may be the best approach for mitigating the risks of military AI.

Endnotes

[1] Artificial intelligence is the field of study devoted to making machines intelligent. Intelligence measures a system's ability to determine the best course of action to achieve its goals in a wide range of environments. Today's AI systems exhibit narrow artificial intelligence, or task-specific intelligence. The field of AI has a number of subdisciplines and methods used to create intelligent behavior, and one of the most prominent is machine learning. For more on definitions of artificial intelligence, see Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge: Cambridge University Press, 2010). For more on definitions of intelligence, see Shane Legg and Marcus Hutter, *A Collection of Definitions of Intelligence*, technical report, Dalle Molle Institute for Artificial Intelligence, June 15, 2007, https://arxiv.org/pdf/0706.3639.pdf. On machine learning, see Tom Michael Mitchell, "The Discipline of Machine Learning," 2006, Carnegie Mellon University, School of Computer Science, Machine Learning Department; and Ben Buchanan and Taylor Miller, *Machine Learning for Policymakers: What It Is and Why It Matters*, Belfer Center, June 2017, https://www.belfercenter.org/sites/default/files/files/publication/MachineLearningforPolicymakers.pdf. For a brief, nontechnical overview of AI and machine learning, see Paul Scharre and Michael C. Horowitz, *Artificial Intelligence: What Every Policymaker Needs to Know, Center for a New American Security*, June 2018, https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know.

[2] Tim Dutton, "An Overview of National AI Strategies," Medium.com, June 28, 2018, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

[3] This paper does not consider second-order effects on international peace and security due to potential political, economic, and societal disruption from AI. These indirect effects of the AI revolution on international security are potentially even more significant, however. For more on these potential second-order effects, see Michael C. Horowitz et al., *Artificial Intelligence and International Security*, Center for a New American Security, July 10, 2018, https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security.

[4] For example, the AI system Pluribus, a joint project by researchers at Carnegie Mellon University and Facebook, achieved superhuman performance in no-limit Texas hold 'em poker without using any machine learning. James Vincent, "Facebook and CMU's 'Superhuman' Poker AI beats Human Pros," the Verge, July 11, 2019, https://www.theverge.com/2019/7/11/20690078/ai-poker-pluribus-facebook-cmu-texas-hold-em-six-player-no-limit.

[5] Some of the most impressive basic research advances in AI come out of a method called deep reinforcement learning, in which machines learn by interacting with the environment. This method has been used to achieve superhuman performance in complex computer games without any human training data or preprogrammed rules of behavior. For more information, see OpenAI, "OpenAI Five," https://openai.com/five/.

[6] There are wide-ranging debates among AI researchers about the future direction of the field. For more on a few of these views, see Rich Sutton, "The Bitter Lesson," incompleteideas.net, March 13, 2019, http://www.incompleteideas.net/IncIdeas/BitterLesson.html; and Rodney Brooks, "A Better Lesson," *Robots, AI, and Other Stuff* (blog), rodneybrooks.com, March 19, 2019, https://rodneybrooks.com/a-better-lesson/.

[7] For some examples of security-related applications of artificial intelligence, see Miles Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*, February 2018, https://maliciousaireport.com/.

[8] Kevin Kelly, "The Three Breakthroughs That Have Finally Unleashed AI on the World," Wired, October 27, 2014, https://www.wired.com/2014/10/future-of-artificial-intelligence/.

[9] For an English-language analysis of Chinese military scholarship on the intelligentization of warfare, see Elsa Kania, *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power*, Center for a New American Security, November 28, 2017, https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power.

[10] Examples of different AI disciplines include neural networks, evolutionary or genetic algorithms, computational game theory, Bayesian statistics, inductive reasoning, fuzzy logic, analogical reasoning, and hand-coded expert knowledge. For more background on AI, see Scharre and Horowitz, *Artificial Intelligence*.

[11] For an overview of the limitations of current narrow AI systems, see Dario Amodei et al., *Concrete Problems in AI Safety*, Cornell University arXiv.org, July 25, 2016, 4, https://arxiv.org/pdf/1606.06565.pdf; Dario Amodei and Jack Clark, "Faulty Reward Functions in the Wild," *OpenAI* (blog), OpenAI, December 21, 2016, https://blog.openai.com/faulty-reward-functions/; and Joel Lehman et al., *The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities*, Cornell University arXiv.org, March 8, 2018, 6, https://arxiv. org/pdf/1803.03453.pdf.

[12] Jim Puzzanghera, "Driver in Tesla Crash Excessively on Autopilot, but Tesla Shares Some Blame, Federal Panel Finds," *Los Angeles Times*, September 12, 2017, http://www.latimes.com/business/la-fi-hy-tesla-autopilot-20170912-story.html; "Driver Errors, Overreliance on Automation, Lack of Safeguards, Led to Fatal Tesla Crash," National Transportation Safety Board Office of Public Affairs, press release, September 12, 2017, https://www.ntsb.gov/news/press-releases/Pages/PR20170912.aspx; "Collision Between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida," NTSB/HAR-17/02/PB2017-102600, National Transportation Safety Board, May 7, 2016, https://www.ntsb.gov/news/events/Documents/2017-HWY16FH018-BMG-abstract.pdf; James Gilboy, "Officials Find Cause of Tesla Autopilot Crash into Fire Truck: Report," The Drive, May 17, 2018, http://www.thedrive.com/news/20912/cause-of-tesla-autopilot-crash-into-fire-truck-cause-determined-report; "Tesla Hit Parked Police Car 'While Using Autopilot,'" BBC, May 30, 2018, https://www.bbc.com/news/technology-44300952; and Raphael Orlove, "This Test Shows Why Tesla Autopilot Crashes Keep Happening," Jalopnik, June 13, 2018, https://jalopnik.com/this-test-shows-why-tesla-autopilot-crashes-keep-happen-1826810902.

[13] Bob van den Hoek, "Can AlphaGo Win Lee Sedol on a Larger Size Board? Say, 4x the Size," Quora, May 14, 2016, https://www.quora.com/Can-AlphaGo-win-Lee-Sedol-on-a-larger-size-board-Say-4x-the-size.

[14] Larry Hardesty, "Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems," MIT News, February 11, 2018, http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212.

[15] Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women," Reuters, October 9, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[16] The number of possible interactions is not technically infinite, but it is a larger number of interactions than could be reasonably calculated.

[17] John Krafcik, "Where the Next 10 Million Miles Will Take Us," Waymo, October 10, 2018, https://medium.com/waymo/where-the-next-10-million-miles-will-take-us-de51bebb67d3.

[18] Anh Nguyen et al., "Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images," Computer Vision and Pattern Recognition, IEEE, 2015; James Vincent, "Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day," the Verge, May 24, 2016; and Nicolas Papernot et al., *Practical Black-Box Attacks against Machine Learning*, Cornell University arXiv.org, March 19, 2017, https://arxiv.org/pdf/1602.02697.pdf.

[19] Paul Scharre, *Robotics on the Battlefield, Part II: The Coming Swarm*, Center for a New American Security, October 15, 2014, https://s3.amazonaws.com/files.cnas.org/documents/CNAS_TheComingSwarm_Scharre.pdf?mtime=20160906082059.

[20] OpenAI, "OpenAI Five."

[21] Krafcik, "Where the Next 10 Million Miles Will Take Us."

[22] Kate Goddard et al., "Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators," *Journal of the American Medical Informatics Association* 19, no. 1 (2012): 121-7, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3240751/.

[23] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W. W. Norton, 2018), 137-145.

[24] Paul Robinette et al., *Overtrust of Robots in Emergency Evacuation Scenarios*, 2016, https://www.cc.gatech.edu/~alanwags/pubs/Robinette-HRI-2016.pdf.

[25] Bernd Schaefer et al., *Forecasting Nuclear War: Stasi/KGB Intelligence Cooperation under Project RYaN*, Wilson

Center, November 13, 2014, https://www.wilsoncenter.org/publication/forecasting-nuclear-war.

[26] Patricia Lewis et al., *Too Close for Comfort: Cases of Near Nuclear Use and Options for Policy*, Royal Institute of International Affairs, London, April 2014, https://www.chathamhouse.org/publications/papers/view/199200; and Scott D. Sagan, *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons* (Princeton, NJ: Princeton University Press, 1993)

[27] David Hoffman, "'I Had a Funny Feeling in My Gut,'" *Washington Post*, February 10, 1999, http://www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/shatter021099b.htm.

[28] Michael Lewis, *Flash Boys: A Wall Street Revolt* (New York: W. W. Norton, 2015), 63, 69, 74, 81.

[29] US Commodity Futures Trading Commission and US Securities and Exchange Commission, "Findings Regarding the Market Events of May 6, 2010," September 30, 2010, 2, http://www.sec.gov/news/studies/2010/marketevents-report.pdf.

[30] D7, "Knightmare: A DevOps Cautionary Tale," *Doug Seven* (blog), April 17, 2014, https://dougseven.com/2014/04/17/knightmare-a-devops-cautionary-tale/.

[31] US Securities and Exchange Commission, "Investor Bulletin: Measures to Address Market Volatility," July 1, 2012, https://www.sec.gov/oiea/investor-alerts-bulletins/investor-alerts-circuitbreakersbulletinhtm.html.

[32] Matt Egan, "Trading Was Halted 1,200 Times Monday," CNN *Money*, August 24, 2015, http://money.cnn.com/2015/08/24/investing/stocks-markets-selloff-circuit-breakers-1200-times/index.html.

[33] Chen Hanghui [陈航辉], "Artificial Intelligence: Disruptively Changing the Rules of the Game" [人工智能：颠覆性改变“游戏规则], China Military Online, March 18, 2016, http://www.81.cn/jskj/2016-03/18/content_6966873_2.htm (Chen Hanghui is affiliated with the Nanjing Army Command College); and John R. Allen and Amir Husain, "On Hyperwar," *Proceedings*, July 2017, https://www.usni.org/magazines/proceedings/2017/july/hyperwar.

[34] For more on the risk of a race to the bottom on safety, see Paul Scharre, "Killer Apps: The Real Dangers of an AI Arms Race," *Foreign Affairs*, (May/June 2019), https://www.foreignaffairs.com/articles/2019-04-16/killer-apps.

[35] For a comprehensive overview of different types of weapons bans, see Scharre, *Army of None*, 331-345.

[36] For controls on other dual-use technologies, see Elisa D. Harris, ed., *Governance of Dual-Use Technologies: Theory and Practice* (Cambridge, MA: American Academy of Arts and Sciences, 2016), http://www.amacad.org/sites/default/files/academy/multimedia/pdfs/publications/researchpapersmonographs/GNF_Dual-Use-Technology.pdf; and Jonathan B. Tucker, ed., *Double-Edged Innovations: Preventing the Misuse of Emerging Biological/Chemical Technologies*, Defense Threat Reduction Agency, July 2010, https://apps.dtic.mil/dtic/tr/fulltext/u2/a556984.pdf.

[37] For example, see United Nations, "Group of Governmental Experts on Transparency and Confidence-Building Measures in Outer Space Activities," July 29, 2013, https://undocs.org/A/68/189.

About the Author

Paul Scharre (pscharre@cnas.org) is a Senior Fellow and Director of the Technology and National Security Program at the Center for a New American Security, an independent, bipartisan, Washington, DC-based think tank.

This working paper was prepared for a workshop, organized by the Stanley Center for Peace and Security, UNODA, and the Stimson Center, on The Militarization of Artificial Intelligence.

Artificial Intelligence and Its Military Implications

China Arms Control and Disarmament Association

July 2019

What Is Artificial Intelligence?

Artificial intelligence (AI) refers to the research and development of the theories, methods, technologies, and application systems for simulating, extending, and expanding human intelligence. One of the main objectives of AI research is to enable machines to do complex tasks that usually require human intelligence to complete. As a branch of computer science, it seeks to understand the essence of intelligence and produce new intelligent machines that respond in a way similar to human intelligence. Such machines may attempt to mimic, augment, or displace human intelligence.

AI can be categorized by certain capabilities. Weak or narrow AI refers to artificial intelligence that can simulate specific intelligent behaviors of human beings, such as recognition, learning, reasoning, and judgment. Strong or artificial general intelligence (AGI) refers to AI that has an autonomous consciousness and innovative ability similar to that of the human brain. To put it differently, weak AI aims to solve specific tasks, such as speech recognition, image recognition, and translation of some specific materials. Strong AI can think, make plans, and solve problems, as well as engage in abstract thinking, understand complex ideas, learn quickly, and learn from experiences, which is near to human intelligence. Artificial superintelligence refers to future AI that will far surpass the human brain in its computing and thinking ability, and is what Oxford University philosopher Nick Bostrom described as "much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills."[1]

There are also people who think AI is hard to define because intelligence is hard to define in the first place. Consensus exists that AI is not natural; it's man-made, yet it can reason and make decisions that take various factors into account. In addition, the term "robot" is not a synonym for AI, even if it is sometimes used that way.[2] Fu Ying, former vice minister of foreign affairs of China, writes, "Our discussion of AI and its impact on international relations and even the global landscape can only be limited to the AI technologies and relevant applications that we know of, which use the three major elements of computing power, algorithms, and data, and are based on big data and deep learning technology."[3] She goes on to suggest that discussion should not focus on possible future AI technologies or capabilities.[4]

AI represents an increasingly multidisciplinary endeavor, and its scope of research goes far beyond computer science to include robotics, language recognition, image recognition, natural-language processing, expert systems, neural networks, machine learning, deep learning, and computer vision. What stands at the core of AI are the often-cited algorithms, computing power, and data, for which the big powers compete.

AI theory and technology are witnessing rapid advances, with increasingly wide application in various domains, such as agriculture, manufacturing, health care, transportation, and even the military. With these advances come social, ethical, and legal implications. AI developers might not always take into account these implications, as that can require proficiency not only in the fields of computer science, psychology, linguistics, and neuroscience but also ethics, law, and philosophy.

Military Application of Artificial Intelligence

Artificial intelligence might affect different aspects of war in unprecedented breadth and depth. For instance, the emergence of predictive maintenance software, intelligent decision-making assistants, autonomous underwater vehicles, or drone clusters could drive a new round of military reform and change the face of war quietly.[5] Fu believes that a state's technological preponderance in AI will quickly become an overwhelming advantage on the battlefield, though it is necessary to understand the military application of artificial intelligence in a holistic way.[6]

On the whole, military applications of artificial intelligence cover two major dimensions. First, AI could be used to improve the performance of traditional and existing weapon systems. Second, AI could assist with or facilitate decision making or make autonomous decisions.

Artificial intelligence might be the most important dual-use technology in the coming decades. Some experts think that AI, as a cutting-edge dual-use technology, has deep and wide application in weapon systems and equipment. Compared with traditional technology, AI-enabled weapon systems would enjoy various advantages, such as having an all-round and all-weather combat capability and a robust survivability on the battlefield, as well as lower cost.[7]

One of the biggest advantages of AI-enabled weapon systems and equipment is response speed, which might far surpass that of the human brain. In a simulated air battle in 2016, an F-15 fighter aircraft operated by the intelligent software Alpha, which was developed by the University of Cincinnati, defeated a human-piloted F-22 fighter aircraft because the intelligent software could react 250 times faster than the human brain.[8]

With the development of AI technologies, intelligent weapon systems that can autonomously identify, lock in on, and strike their targets are on the rise and can perform simple decision-making orders in place of humans.[9] However, these systems possess a low level of intelligence, and the mode of autonomous engagement is usually the last option. But when intelligent technologies progress—such as sensor technology and new algorithm and big-data technology—the autonomy of weapon systems will experience great improvement, and the autonomous confrontation between weapon systems will become commonplace. In certain areas of warfare, such as cyberspace and the electromagnetic spectrum, humans can only rely on intelligent weapon systems for autonomous confrontation. When hypersonic weapons and cluster operations arise, war will enter the era of flash wars[10] during which the autonomous fighting between intelligent systems might be the only way out.

Moreover, AI technologies could be used for intelligent situational awareness and information processing on the battlefield and in unmanned military platforms such as certain aerial vehicles and remote-controlled vehicles. Intelligent command-and-control systems developed by militaries could aid decision making and improve the capacity for intelligent assessment. For instance, the US Cyber Command is attempting to strengthen its cyber offensive and defensive capabilities, with a focus on developing intelligent information systems for analyzing cyberintrusion based on cloud computing, big-data analysis, and other technologies. This approach aims to automatically analyze the source of cyberintrusion, the level of damage to networks, and the data-recovery ability.

The military application of AI would also exert a great impact on military organization and combat philosophy, with the potential for fundamentally changing the future of warfare.[11] For example, the combined application of precision-strike ammunition, unmanned equipment, and network information systems has brought about new intelligent combat theories, such as combat cloud and swarm tactics.[12] With its increasingly extensive application in the military field, AI is becoming an important enabler of military reform, giving birth to new patterns of war and changing the inherent mechanism of winning a war. In July 2016, the US Marine Corps tested the modular advanced armed robot system, which uses sensors and cameras to control gun-toting robots based on AI. Israeli tech firm General Robotics Ltd. has developed DOGO, which *Defense News* described as the "world's first inherently armed tactical combat robot."[13] DOGO is similar to a land-based combat drone. This robot could "revolutionize the way commando units and SWAT teams conduct counterterrorism operations around the world, which is precisely what it was created to do."[14]

AI can enhance the effectiveness of war prediction in at least two ways. One is by calculating and predicting war outcomes more accurately. With the support of advanced algorithms and supercomputing capabilities, the calculative and predictive results of AI systems might be more accurate than in the past. The other is by testing and optimizing war plans more effectively with the help of war-game systems integrated with AI.[15] For instance, an AI-integrated war-game system can conduct human-machine confrontation, which contributes to finding possible problems and weak points. In addition, such war-game systems can also be used to develop machine-machine confrontation and improve their efficiency.

AI-enabled decision aids can also free up human capacity, allowing humans to focus on major decisions and key tasks in future wars. It is noteworthy, however, that while AI enjoys wide application in the military field, human soldiers remain the ultimate decision makers for when to move into and out of the chain of operations and to take necessary intervening measures. The biggest challenge for the development of human-machine collaborative technology is ensuring humans take over at any time.[16]

Fu also points out "there is still a great deal of uncertainty regarding the impact of AI on military affairs, both in terms of the extent and forms of impact." Experts on strategic studies still tend to analyze their impact on operations from a single perspective. Fu argues that without a holistic understanding of the military applications of AI, the proposed responses could become "a new Maginot line," expensive and useless.[17]

Emerging Issues in the Military Application of AI

Just like other emerging technologies, AI is a double-edged sword. In particular, along with the increasingly wide military application of AI, some issues have emerged and aroused concern across the world. Bostrom, in a report on global disaster risks, argued that

AI is more serious than nuclear weapons and environmental disasters.[18]

AI Arms Racing and Arms Control

There is concern about an AI arms race. The late British physicist Stephen Hawking said, "Governments seem to be engaged in an AI arms race, designing planes and weapons with intelligent technologies."[19] The competition for global leadership in AI has been under way for some time. In 2017 and 2018, Canada, Japan, Singapore, China, the United Arab Emirates, Finland, Denmark, France, the United Kingdom, the European Commission, South Korea, India, and others all released strategies to promote AI application and development. These strategies focus on different areas, as AI policy researcher Tim Dutton has written: "scientific research, talent development, skills and education, public and private sector adoption, ethics and inclusion, standards and regulations, and data and digital infrastructure."[20] So, it seems that nations will "spar" over AI through competition in research, investment, and talent.[21]

Kenneth Payne of King's College London wrote in *Survival* that "the idea of arms control for AI remains in its infancy" because "the advantages of possessing weaponized AI are likely to be profound and because there will be persistent problems in defining technologies and monitoring compliance."[22] Military application of AI is often compared to the use of electricity.[23] As with using electricity, no country could be banned from using AI. Just as with the arms race between the United States and the Soviet Union during the Cold War, "an algorithm race between AI powers is likely to emerge in the future."[24] But unlike the arms-control agreements reached between the United States and the Soviet Union at that time, such a consensus on an algorithm-control agreement is unlikely to materialize, given the current state of major power relations.

Ethics

In recent years, along with the development of AI research and industry, some pertinent ethical and social problems have become increasingly prominent. They include security risks, privacy, algorithmic discrimination, industrial impact, unemployment, widening income distribution differences, responsibility sharing, regulatory problems, and impact on human moral and ethical values.

Zeng Yi, a research fellow from the China Academy of Sciences, commented that as a result of design flaws, many AI models at this stage are more concerned with how to get the maximum computing reward but ignore the potential risks to the environment and society. "The vast majority of AI today does not have a concept of self and cannot distinguish between self and others. Human experience, the speculation of external things, is based on one's own experience," he said.[25]

AI systems cannot really understand human values. This is one of the biggest challenges in AI. So it is important for AI ethics studies to consider how to make a machine to self-learn human values and avoid risk. Also, the ethical code of AI should be a topic in the dialogue among various countries and organizations.

In the military field, there are similar ethical problems, in particular those concerning human dignity in the face of autonomous weapons systems. Therefore, research on AI ethics and security is needed and should integrate the efforts of technology and society to ensure that AI development remains beneficial to human beings and nature. Of course, "technological developments will raise new requirements for ethical codes," Zeng said. "However, given the differences in culture and places, it is not only difficult to implement the proposal of unified guidelines, but also unnecessary. Therefore, it is very important to coordinate the ethical standards among different countries and organizations."[26]

Legal Governance

So far there have been more than 40 proposals for AI ethics guidelines issued by national and regional governments, nongovernmental organizations, research institutions, and industries. For instance, in April 2015, the International Committee of the Red Cross published advisory guidance on the use of autonomous weapons.[27] But the various guidelines employ different perspectives on specific issues, and "none of them covered more than 65 percent of the other proposals," according to Zeng.[28]

Also, customary and formal international law remains in flux. In April 2019, the European Commission released a code of ethics for AI and announced the launch of a trial phase of the code, inviting companies and research institutions to test it. On May 25, 2019, the Beijing Academy of AI released the Beijing AI Principles.

In terms of research and development, AI should be subject to the overall interests of humankind, and the design should be ethical; in order to avoid misuse and abuse, AI should ensure that stakeholders have full knowledge and consent of the impact on their rights and interests; in terms of governance, we should be tolerant and cautious about replacing human work with AI.[29] The Tsinghua Center for International Strategy and Security in Beijing proposed six principles for AI related to well-being, security, sharing, peace, rule of law, and cooperation. It also pointed out that these principles are still vague and abstract and that it takes time to refine and discuss them with experts from other countries to find the greatest common divisor.[30] From these proposals, the necessity and urgency for AI governance, especially its military applications, can be detected.

When autonomous weapon systems (AWS) and AI are employed in warfare, the consequences cannot be overestimated. A legal framework to govern the military use of AI is urgently needed. Several issues deserve more discussion:[31]

– AWS must be defined, including clarifying the differences in the autonomy of mines, unmanned aerial vehicles, and missiles.

– There is a need to explore pragmatic principles governing autonomous weapons and AI. For instance:

- Should a commander be asked to activate a machine because it can respond faster than a human being?
- What preventive measures should be taken?
- What is due legal deliberation?
- How can offenders be held accountable for intentional violations of international law? Is malfeasance a crime?
- How does one tell if an attack is imminent?
- How can human judgment and human control over the machine be guaranteed?

– There is a need to discuss the legal threshold for the use of force, including self-defense and countermeasures.

– There is a need to protect civilians from autonomous weapons. For instance, after years of deploying drones in Afghanistan, the United States might have learned lessons and gained experiences in preventing civilian casualties.

As the issue of AI ethics now draws wide attention, there are opportunities to explore how to apply international laws to AI technology. In the military sense, AI poses a number of problems for international law, which need further clarification and exploration. For instance:

– Will the principles of international humanitarian law and the law of war be applicable to AI weapons? For example, the principle of distinction between military and civilian targets, the principle of proportionality that prohibits excessive attacks, the principle of military necessity, and restrictions on means of combat.

– Is there a need for specific rules for AI weapons?

– How should belligerents distinguish combatants from noncombatants in intelligent warfare?

– Should war robots be humanely treated?

– Should AI weapons be accountable for the damage they cause? If not, then should the manufacturer or the user of the weapon be held accountable?

– When AI weapons violate the principle of state sovereignty, will their actions trigger state responsibility?

Of course, as with nuclear weapons and many other military technologies, "norms will likely follow technology, with law materializing still later."[32]

International Cooperation

Artificial intelligence can significantly improve global productivity and promote world economic development. It can also widen the gap between developed economies and developing countries, alter global supply chains, and change the structure of employment and production. Its military application also draws much attention from both theorists and practitioners.

In his congratulatory message to the 2018 World Conference on Artificial Intelligence on September 17, 2018, Chinese President Xi Jinping said:

"A new generation of artificial intelligence is booming around the world, injecting new momentum into economic and social development and profoundly changing people's way of life. To grasp this development opportunity and deal with the new issues raised by artificial intelligence in law, security, employment, moral ethics, and government, governance requires countries to deepen cooperation and discuss it together.

"China is ready to work with other countries in the field of artificial intelligence to promote development, protect security, and share the benefits. China is committed to high-quality development. The development and application of artificial intelligence will effectively improve the level of intelligence of economic and social development, effectively enhance public services and urban management capabilities. China is willing to exchange and cooperate with other countries in technology exchange, data sharing and application market to share opportunities for the development of digital economy."[33]

International law applies to cyberspace as well as to AI. In cyberspace, experts from different fields communicate with each other, as should be the case with AI, which will help the understanding of how the law applies to AI. Countries can use confidence-building measures and exercise self-restraint. Specific guidelines are often derived from practice, but possible scenarios and security concerns can also be discussed, with a view to furthering international cooperation, making AI a force for good, and bringing AI potential into full play while avoiding possible negative effects.

Conclusion

In 2018, UN Secretary-General Antonio Guterres issued an important document, *Securing Our Common Future: An Agenda for Disarmament*, which outlined a comprehensive disarmament agenda and relevant action plans.[34] He also emphasized the importance of dealing with the emerging means and methods of warfare, including keeping weapons and AI in human control. In the future, under the leadership of the United Nations and active participation and cooperation of states, humanitarian actors, civil society, and the private sectors, the international community needs to explore effective governance and risk mitigation of the AI application to enhance sustainable peace and security for all.

Endnotes

[1] Nick Bostrom, "How Long Before Superintelligence?," *International Journal of Future Studies* 2 (1998), https://www.nickbostrom.com/superintelligence.html.

[2] Ben Dickson, "What Is Narrow, General and Super Artificial Intelligence," TechTalks, May 12, 2017, https://bdtechtalks.com/2017/05/12/what-is-narrow-general-and-super-artificial-intelligence/.

[3] Fu Ying, "A Preliminary Analysis of the Impact of AI on International Relations," *Quarterly Journal of International Politics* (April 10, 2019), https://pit.ifeng.com/c/7lkmTsTwMD2, trans. Brian Tse and Jeffrey Ding, https://docs.google.com/document/d/1zYU-29n9oIKcMZXM1j3_DMecY77LAs-bdeK5m6CbQ5A/edit?pli=1#.

[4] Fu Ying, *Impact of AI upon International Relations*, International Strategy and Security Studies Report, No. 1, 2019, Center for International Strategy and Security, Tsinghua University, 3.

[5] Chen Hanghui, "Artificial Intelligence: How Would AI Subvert Future War," *China National Defense News*, January 2, 2018, http://www.mod.gov.cn/jmsd/2018-01/02/content_4801253.htm.

[6] Fu, *Impact of AI*, 8.

[7] Yang Xiaonan and Ma Mingfei, "Artificial Intelligence Becomes the Booster of National Defense," *PLA Daily*, July 11, 2018, http://www.mod.gov.cn/jmsd/2018-07/11/content_4819011.htm.

[8] Chen, "Artificial Intelligence."

[9] For example, the US Aegis system, the Israeli Iron Dome system, the Russian Afghanit active protection system, and the French Shark system.

[10] Cluster operation is the centralized deployment of numerous intelligent weapons to attack targets from multiple directions, which can achieve the effect of "quantity is quality" and is perceived as AI-era attrition war. The potential use of drones might serve as an example of this mode of future operations on the battlefield. See Chen, "Artificial Intelligence"; U. Franke, "Flash Wars: Where Could an Autonomous Weapons Revolution Lead Us?," European Council on Foreign Relations, November 22, 2018, https://www.ecfr.eu/article/Flash_Wars_Where_could_an_autonomous_weapons_revolution_lead_us.

[11] Yang and Ma, "Artificial Intelligence."

[12] "The combat cloud conveys a system in which data is pooled and is available from this via a number of different means. The essence of the 'cloud' notion in combat cloud is that a user is not dependent upon information being pushed to them via a specific means; they are connected to the cloud via whatever means they have at their disposal, and can pull data they are authorized to see as and when necessary." Chris McInnes, "The Combat Cloud," Defense.info, June 30, 2018, https://defense.info/williams-foundation/2018/06/the-combat-cloud/.

[13] B. Opall-Rome, "Introducing: Israeli 12-Kilo Killer Robot," *Defense News*, May 8, 2016, https://www.defensenews.com/global/mideast-africa/2016/05/08/introducing-israeli-12-kilo-killer-robot.

[14] Adam Linehan, "Meet DOGO: The Cute Little Robot out for Terrorist Blood," Task & Purpose, May 11, 2016, https://taskandpurpose.com/meet-dogo-cute-little-robot-terrorist-blood.

[15] Chen, "Artificial Intelligence."

[16] Ibid.

[17] Fu, "Preliminary Analysis."

[18] See N. Bostrom, *SuperIntelligence: Paths, Dangers, and Strategies* (Oxford: Oxford University Press, 2014).

[19] Stephen Hawking, interview by Larry King, *Larry King Now*, ORA TV, June 25, 2016, http://www.ora.tv/larrykingnow/2016/6/25/larry-kings-exclusive-conversation-with-stephen-hawking.

[20] Tim Dutton, "An Overview of National AI Strategies," Medium, June 28, 2018, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

[21] Price Waterhouse Cooper, "Nations Will Spar over AI," https://www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions/ai-arms-race.html.

[22] Kenneth Payne, "Artificial Intelligence: A Revolution in Strategic Affairs?," *Survival* 60, no. 5 (October–November 2018): 19.

[23] See Michael Horowitz, Elsa Kania, Gregory C. Allen, and Paul Scharre, *Strategic Competition in an Era of Artificial Intelligence*, Center for a New American Security, July 25, 2018, https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence.

[24] Fu, *Impact of AI*, 15.

[25] Ren Fang Yan, "Zeng Yi: Multi-Party Artificial Intelligence Ethics Must be Coordinated," ScienceNet, April 21, 2019. http://news.sciencenet.cn/htmlnews/2019/4/425476.shtm.

[26] Ibid.

[27] International Committee of the Red Cross, "Autonomous Weapon Systems: Is It Morally Acceptable for a

Machine to Make Life and Death Decisions?," statement of the International Committee of the Red Cross, April 13, 2015, https://www.icrc.org/en/document/lethal-autonomousweapons-systems-LAWS.

[28] Ren, "Zeng Yi."

[29] Beijing Academcy of Artificial Intelligence, "Beijing AI Principles," May 28, 2019, https://www.baai.ac.cn/news/beijing-ai-principles-en.html.

[30] Fu, *Impact of AI*, 20.

[31] See United Nations Institute for Disarmament Research (UNIDIR), *The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches*, UNIDIR Resource, No. 6., 2017, https://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-concerns-characteristics-and-definitional-approaches-en-689.pdf.

[32] Payne, "Artificial Intelligence," 19.

[33] Xi Jinping, "Congratulatory Letter from Xi Jinping to the 2018 World Artificial Intelligence Conference," September 17, 2018, http://www.mod.gov.cn/shouye/2018-09/17/content_4825134.htm.

[34] UN Office for Disarmament Affairs, *Securing Our Common Future: An Agenda for Disarmament*, 2018, https://www.un.org/disarmament/sg-agenda/en/.

About the Author

The China Arms Control and Disarmament Association (CACDA) is a nonprofit, nongovernmental organization (NGO) founded in August 2001 in Beijing. CACDA undertakes the organization and promotion of academic research and nongovernmental activities at home and abroad in the areas of arms control, disarmament, and nonproliferation so as to facilitate international endeavors for world peace and security. The association was granted NGO in Special Consultative Status with the Economic and Social Council of the United Nations in 2005.

This working paper was prepared for a workshop, organized by the Stanley Center for Peace and Security, UNODA, and the Stimson Center, on The Militarization of Artificial Intelligence.

Discussion Paper | A Russian Perspective

Militarization of AI

Vadim Kozyulin | PRI Center (Russian Center for Policy Research)

July 2019

The Robotization Program in Russia

The term "artificial intelligence" manifested itself loudly at the state level in Russia in 2017. "Artificial intelligence is not only the future of Russia, it is the future of all mankind. There are enormous opportunities and threats that are difficult to predict today. The one who becomes a leader in this sphere will be the ruler of the world," President Vladimir Putin said on September 1, 2017.[1] The topic immediately became popular in the Russian media scene, businessmen and government officials discussed prospects of AI development in Russia at various forums, and a wave of information-technology forums for specialists and startups swept across the country. The Russian government has released a national strategy for artificial intelligence.[2] That decree directed the government to formulate and approve a Federal Project on Artificial Intelligence as part of the national program called Digital Economy of the Russian Federation. Up to 90 billion rubles ($1.4 billion) will be spent for these purposes in six years.[3]

The Russian military began to use the term "artificial intelligence" around 2017, when the Ministry of Economic Development held the roundtable "Artificial Intelligence" at the Military-Technical Forum ARMY-2017.[4] Since then, no conference held by the Russian Defense Ministry (MoD) has avoided this topic.

Until 2017, what is now associated with military artificial intelligence was associated with robotics in Russia. The Military Encyclopedic Dictionary on the official website of the Russian Ministry of Defense cites the concept of "combat robot" as a "multi-functional technical device with anthropomorphic (humanlike) behavior, partially or fully performing functions of a person in executing certain combat missions. It includes a sensor system (sensors) for acquiring information, a control system, and actuation devices."[5]

The Russian military divides combat robots into three generations:

- First-generation robots have software and remote control that can only function in an organized environment.
- Second-generation robots are adaptive to changes in their environment, having a kind of sensory organs and an ability to function in a random environment.
- Third-generation robots are smart robots equipped with an AI-based control system. So far, such robots are only available as laboratory models.

Unmanned tanks and torpedo boats, robot soldiers, and others that are used to support combat activity of troops in conditions adverse to humans should be regarded as the simplest combat robots.

In 2000, the Russian MoD adopted an integrated target program, Robotization of Weapons and Military Equipment—2015. The program allowed for successful research and development, with experimental mock-up models of ground-based robotic systems produced and tested. However, development and engineering never started, which led to the suspension of research and development in ground-based military robotics.[6]

In September 2015, the MoD started the program Creation of Advanced Military Robotics for 2025. The program prioritized "designing of unmanned vehicles in the form of robotic systems and complexes for military use in various environments of application." The General Staff of the Russian Armed Forces developed a plan for the use of robotic systems for military purposes until 2030 and approved the general technical requirements for military ground robotic systems. According to this plan, about 30 percent of Russian military equipment should be remotely controlled by 2025.[7]

Following the measures taken by the MoD on December 16, 2015, Putin signed a decree establishing the National Center for

Development of Technologies and Basic Elements of Robotics, which entrusted the Foundation for Advanced Research to furnish the center's activities.

In December 2016, the Russian government adopted the *Strategy of Scientific and Technological Development of the Russian Federation*, which named as priorities "the transition to advanced digital, intelligent manufacturing technologies, robotic systems, new materials and methods of construction, development of systems for big data processing, machine learning and artificial intelligence."[8]

Finally, Russia finished shaping the network of organizations responsible for military blueprints. That network or structure currently consists of:

- The MoD Commission for the Development of Robotic Systems for Military Purposes: Headed personally by Russian Defense Minister Sergei Shoigu, the commission develops a common model and procedures for designing robotic systems, reducing their types, and unification and coordination between various departments.

- The Main Department for Research and Technological Support of Advanced Technologies (GUNID): Part of the Russian MoD, GUNID is the prime contract specifier of military robots, and it also develops unified ideology and engineering procedures.

- The Main Research and Testing Robotics Centre of the Russian Ministry of Defense (MRTRC): The center is one of the most secretive military organizations in the country and rarely reports its achievements. It is known that the center creates the Russian marine information and measurement network intended for regular observations in the Arctic. According to the head of MRTRC, Sergey Popov, "the trend 'smart, small, many, and inexpensive' has gradually gained a realistic status, which is proven by specific achievements in modern robotics. "[9]

- The Advanced Research Foundation: According to Putin, "the Foundation's projects are designed to play a decisive role in the development of key elements of the new generation of weapons, military, and special equipment. They should become the basis of the national weapons system at the turn of 2025–2030 both for the Army and Navy, and for a number of other industries and law enforcement agencies."[10]

- The Military Innovative Technopolis "ERA": By 2020, twelve scientific companies of the MoD with approximately 600 personnel will work at this new research campus.[11] The technopolis will conduct research and development in priority areas, including information and telecommunication systems; automated control systems; robotic systems; artificial intelligence systems; computer modeling; information security; technical vision and pattern recognition; nanotechnology and nanomaterials; computer science and computer technology; energy, technologies, devices, and life-support machines; and bioengineered, biosynthetic, and biosensor technologies.[12] More than 80 leading Russian research and industrial enterprises plan to open laboratories and engineering units in the technopolis.

Russian manufacturers of military equipment follow their colleagues from the United States, Israel, and South Korea in designing autonomous combat robots and extending their capabilities, like reconnaissance and surveillance, patrolling, fire support, protection of objects and the breaching of penetrations in the barriers, delivery of ammunition and medevac, installation of minefields and demining, setting smoke screens, and even mobile audio propaganda. The Russian military has already learned to increase the effectiveness of combat systems through use of AI technologies. The Central Research Institute of Aerospace Defense Forces and the Research Institute of Electronic Warfare conducted research that demonstrated a twofold increase in the efficiency of air and missile defense when working hand in glove with early warning systems.[13]

Modern neural networks allow autonomous weapons systems such as unmanned aerial and ground combat vehicles to not only come to independent decisions but also to adapt to the changing environment. Russian manufacturer Kalashnikov has designed a ground-based "battle module based on neural networks" that is able to "gain targets and make decisions" without the operator's engagement. The BAS-01G Soratnik combat automated system, armed with a PKTM tank machine gun and Kornet-EM antitank missiles, utilizes "lessons learned" when performing new combat missions.[14]

In February 2019, Kalashnikov was the first Russian manufacturer to develop a loitering munition—the KUB-BLA. Six months later, Kalashnikov revealed a new high-precision, unmanned complex called the ZALA Lancet. Such weapons are known as kamikaze drones because of their capacity to independently detect and attack targets. The website Kalashnikov Media defines the Lancet as an intelligent, multitasking weapon that can independently acquire an assigned target and attack it.[15]

The Russian defense industry and the Russian Armed Forces have made a concerted push to close the technological gap that has formed over what they perceive as two decades of inattention and underfunding. A view persists, including among arms manufacturers, that Russian engineers have to redouble their efforts to not lag behind the leading states in such areas as drone "swarming." This issue is couched as a technical task with no focus on moral or philosophical dimensions. National security interests and technological rivalry provide the Russian military with a reason to postpone moral considerations, which could further intensify arms race dynamics.

Consequently, Russian civil society does not pay much attention to the problem of human control over lethal autonomous weapons systems, and its understanding is limited about the debate that takes place through the Convention on Certain Conventional Weapons on the legality of such systems in military conflicts.

While the development of AI opens up new opportunities for the military, it is also fraught with new risks. The opportunities and risks have yet to be fully comprehended.

Combining emerging technologies with existing military resources increases efficiency in new ways. However, the speed of technological changes amplifies the arms competition that is present in rival militaries.

Future Threats to Crisis and Strategic Stability

In June 1990, the Washington Summit Joint USSR-US Statement on Future Negotiations on Nuclear and Space Arms and Further Enhancing Strategic Stability defined strategic stability as a state of relations between the two powers where, even in a crisis, neither side has serious opportunity and incentive for a nuclear first strike.[16] Arms race stability depends on whether there are incentives to build up a country's strategic potential. The principles of strategic stability formalized in the 1990 statement were considered as guidance for arms control.

This view on strategic stability obviously only takes into account the nuclear capabilities of the two leading nuclear powers—the United States and Russia—and leaves arsenals of other countries out of the formula. There are attempts to define strategic stability considering multilateral military capabilities today. Some offer new concepts of "multilateral" and even "cross-domain strategic stability."

Military artificial intelligence could undermine the foundation of strategic stability in any concept, including the classic American-Russian version. Some senior Pentagon strategists have already made statements that the most cutting-edge technologies and systems—especially from the fields of robotics, autonomous systems, miniaturization, big data, and advanced manufacturing—can provide military dominance.[17] Several technologies have the potential to impact global security today.

US missile defense is one of the most technically complex military projects in history. Automated launch and targeting is a key capability within systems like the Aegis Ballistic Missile Defense System.[18] Although the Russian military has repeatedly stated that US missile defense does not pose a threat to the Russian nuclear triad, doubts grow stronger as the number of deployed US antimissile systems and their capacities improve.[19] It incentivizes Russia to look for new ways to guarantee its nuclear deterrence. Russia responded to this threat at its borders by accelerating the development of a variety of innovative weapons systems, primarily hypersonic missiles. The competition between these heavily automated defense and attack systems further undermines strategic stability.

New, potentially AI-based, long-range, antiship missiles—such as the Russian 3M-55 Onyx[20] or the US LRASM—represent a class of autonomous ship killers. Using AI and datalinks, these missiles make decisions on their own, conducting a coordinated attack on an enemy fleet. They size up the enemy fleet, locate the target, and calculate the desired point of impact.[21]

Interconnection of space tracking and surveillance system with command and control, battle management, and communications by means of AI-based programs opens new possibilities for interception of ballistic missiles. The deputy chief of the Russian General Staff, Viktor Poznikhir, underlined that US radars can monitor flights of Russian ICBM warheads. In addition, US missile defense poses a threat to almost all Russian low-orbit satellites within the reach of the system.[22]

In the near future, fleets of unmanned robotic systems could flood the oceans in order to detect and trace ships and submarines. The prototype of such drones is the Sea Hunter, recently deployed by the US Navy. Should the autonomous marine hunters become a component of antisubmarine warfare, the global oceans would become more transparent, and the invulnerability of nuclear submarines would be questioned, as would their ability to provide strategic deterrence.[23]

Machine learning and autonomy open up the possibility of using nuclear weapons—like the B61-12 low-yield, high-precision nuclear bombs—to accomplish tactical tasks. If AI applications result in improved targeting and coordination, then they could enable precision strikes with low-yield weapons to destroy key command, control, and communication assets—including nuclear force and space monitoring systems—without the use of high-yield nuclear weapons. Similar capabilities could enable the use of nonstrategic, precision weapons to execute strategic operations. New types of nuclear weapons and deployment of ballistic missiles with nonnuclear warheads make arsenals unpredictable and affect strategic stability.

Swarms of unmanned vehicles could open a new page in the history of noncontact warfare without the battlefield presence of human combatants. Inexpensive, expendable drones could quickly map an adversary's combat systems, targeting and destroying key components of its C2 and defense systems with relatively little cost. Threats of this kind require militaries to closely monitor new technologies and to develop new means of defense, which ultimately fuel arms race dynamics.

The inexpensive escort drones for combat aircraft could serve as carriers of weapons or "consumables" in case of combat and would significantly increase combat effectiveness.[24] More-sophisticated air defense and early warning capabilities would shape a response to such a threat, ultimately leading to militarization of the parties at a higher technological level. The introduction of technology for air refueling of attack drones, as well as deployment of unmanned refueling aircraft, could increase force projection range and reduce risks to pilots and aircraft carriers. It would increase tension in regional conflicts and incentivize development of new means to neutralize the remote threat of this kind.

Technological progress makes military strategists think about the permanent removal of pilots from cockpits where a human becomes a hindrance.[25] Preliminary sketches of the prospective sixth generation fighter show an autonomous vehicle capable of fully robotic flight at hypersonic speed, with improved stealth technology across the full electromagnetic spectrum, protected by laser systems and equipped with powerful electronic warfare.

Great prospects exist for military AI in outer space. Unmanned reusable space aircraft (like the Boeing X-37B Orbital Test Vehicle, XS-1 Spaceplane, or X-43A Hypersonic) could shape a new model of space confrontation. Even without weapons on board, these vehicles will cause concern for many militaries and incentivize the development of new defense systems to protect against unknown threats, primarily against spacecraft capable of disabling satellites.[26] For example, France reportedly plans to equip Syracuse telecommunications satellites with cameras and self-defense devices, such as blinding lasers or machine guns for breaking solar panels of an approaching satellite.[27]

Outsourcing of Command, Control, Communication, Intelligence, Surveillance, and Reconnaissance to AI

AI learns to cope with increasingly complex tasks as it accumulates experience and absorbs new technologies. It is already able to not only collect but also to analyze intelligence.

In 2014, the National Defense Operations Center of the Russian Federation was inaugurated in Moscow. It is designed to collect, summarize, and analyze information on the military and political situation in the world and conduct centralized control of the Russian Armed Forces.[28] Based on requests, its software and hardware system monitors and analyzes information from open sources, simulates forecasts of key world events, and prepares recommendations in an automated mode.[29]

The United States and the United Kingdom have similar centers, while other countries are seeking the capability. As the amount of information to be processed increases and the dynamics of events accelerate, the temptation grows to place AI in charge of not only developing recommendations but also drafting and choosing the right scenario for crisis situations. Participation of AI in assessing the situation and responding to threats would increase the risks of unintended military conflict.

Currently, the Russian military is rather reluctant to introduce AI into military affairs. The president of the Russian Academy of Missile and Artillery Sciences, Professor Vasily Burenok, makes reference to the risks of design errors, physical damage, and hostile software impact on AI-based systems. Burenok believes that it is almost impossible to create algorithms suitable for all combat scenarios. In his opinion, AI might only be used in military operations for preparing initial information for decision making by the commander.[30]

In the meantime, the improvement of software, the advancement of learning programs, and the time pressure on decision makers in crisis situations could incentivize using military artificial intelligence. The development of military technologies dramatically reduces the time available to decision makers for assessing threats. The transfer of this function to the AI-based machine may not be a military or political decision but rather a purely technical unwitting one.

Commercial Companies in the Military Domain

Commercial projects based on AI are increasingly being used in the military domain. For example, unmanned vehicle control programs and visual recognition algorithms are equally suitable for commercial vehicles and combat autonomous systems. Military contractors create a new market and new technical horizons for the civilian sector. For their part, commercial companies can offer projects that are interesting to the military, including analysis of satellite images, internet traffic, social network data, global media, air and sea traffic, and even bank transfers.

Russian Defense Minister Shoigu said at the conference "Artificial Intelligence: Problems and Solutions" that military and civil scientists must develop organizational proposals that should be aimed at collaboration between the scientific community, the government, and industrial enterprises.[31]

Meanwhile, the share of private science in Russia is rather small. State research institutes account for 72 percent of domestic research and development (R&D) expenses, and they employ about 80 percent of all Russian researchers.[32] *The Strategy of Scientific and Technological Development of the Russian Federation*, approved by presidential decree on December 1, 2016, admits that "there is a problem of immunity of the economy and society to innovations, which prevents the practical application of the R&D results (the share of innovative products in the total output is only 8-9%; investments in intangible assets in Russia are 3 to 10 times lower than in the leading countries; the share of Russian high-tech products in world exports is about 0.4%). There is virtually no transfer of knowledge and technology between the defense and civilian sectors of the economy, which hinders the development and use of dual-use technologies."[33]

Unlike in Russia, where the state is making its first attempts to attract commercial companies to military contracts, it has become a trend in China, the United States, and other countries. For example, the United States has expanded what President Dwight Eisenhower called in his farewell address a "military-industrial complex."[34] American companies such as Amazon, Microsoft, IBM, Oracle, and Google are now involved in defense projects. Tech giants Google, Apple, Salesforce, and IBM realize the prospects of systems with artificial intelligence and seek to acquire companies engaged in AI.[35] Military AI represents a large market for advanced companies, and the market is undergoing radical changes. "We used to talk about numbers of tanks, planes, ships, troops, but now we have to add components like data centers, supercomputers, simulation speed, and recognition speed to the equation," says Holger Mueller, principal analyst and vice president at Constellation Research Inc.[36] The US Defense Advanced Research Projects Agency plans to invest up to $2 billion in artificial intelligence systems in 2019–2024.[37]

The US Air Force believes commercial space capabilities can improve nuclear triad operations. "Whether it's Silicon Valley or commercial space, there's unlimited opportunities ahead right now for us in terms of how we think differently on things like nuclear command and control," says Air Force Chief of Staff Gen. David Goldfein.[38]

Among the negative side effects of involving civilian companies in military projects, one can mention that the use of civilian infrastructure for military purposes makes civilian objects—such as satellites, transport infrastructure, production facilities, and design bureaus—justifiable targets in the event of a military conflict. According to the United Nations, the United States has more than 1,900 satellites in orbit around the Earth.[39] The Pentagon is actively using their data for military purposes, which raises serious concerns.

The Risk of the Arms Race

In light of the unfolding global race of military technologies, Russia faces a difficult choice between the upcoming reduction in military spending and the need to maintain technological parity with leading states. According to the Stockholm International Peace Research Institute, Russian defense expenditures decreased 19 percent in 2017 compared to 2016, and an additional 3.5 percent in 2018.[40]

On July 4, 2019, Putin, in an interview with the Italian newspaper *Correra Della Serra*, gave rather divergent interpretations of the Russian military development plans:

"Compare the Russian spending on defense—about 48 billion dollars—and the US military budget, which is more than 700 billion dollars. Where is the arms race? We're not going to get involved in it. But we also have to ensure our security. That is why we are bound to develop the latest weapons and equipment in response to the US increase in military spending and its clearly destructive actions."[41]

However, the Russian war chest for military inventions is modest. According to official data that Russia submitted to the United Nations within the Military Expenditures Report, its research and development spending amounted to 687 million rubles in fiscal year 2016 (about $11 million). By comparison, the United States spent $69.04 billion on military research and development in 2019.[42] The Pentagon spends much more on R&D than the Russian MoD, even taking into account that some items of expenditure are not published. Secret and top-secret expenses amounted to 3 trillion rubles in the Russian draft federal budget in 2019. This was 16.9 percent of the expenses for the fiscal year, or 2.9 percent of the gross domestic product. Experts at the Russian Presidential Academy of National Economy and Public Administration and Gaidar Institute estimated in 2018 that the share of classified budget expenditures will grow from 11.6 percent in 2012 to 20.6 percent in 2021.[43]

Russia cannot afford to save on military AI because today's savings could result in a catastrophic strategic loss tomorrow. The famously forward-thinking CEO of Salesforce, Marc Benioff, explained, "Today, only a few countries and a few companies have the very best AI. Those who have the best AI will be smarter, healthier, richer, and their warfare will be significantly more advanced. ... Those without AI are going to be weaker and poorer, less educated and sicker."[44]

Asymmetric Response as a Refuge for Laggards

Deployment of new weapons by technologically advanced powers has always made a strong impression on rivals. Unfortunately, it rarely produces the intended effect that deploying powers expect. Any new, potentially existential threat forces rivals to increase expenditures and find more-destructive ways to ensure security. This is how new missile and nuclear powers were born, and how exotic and dangerous asymmetric responses emerged. In the 21st century, many countries will find themselves lagging behind in military AI. We cannot predict what AI technology may be the most suitable for the implementation of such a "loser strategy." History shows how the world lived with an asymmetric gun to the head throughout the latter half of the 20th century. A military technology race is dangerous in that the perception of falling behind often provokes an asymmetric response.

On November 21, 2017, the head of the Russian Federation Council Committee on Defense and Security, Viktor Bondarev, confirmed deployment of an intercontinental range, nuclear armed, liquid fueled, ballistic SS-N-23 (Skiff) missile emplaced on the ocean floor.[45] In February 2019, Putin announced the end of a key stage of testing of the nuclear powered, unmanned, underwater vehicle *Poseidon*, which can deliver a conventional and a thermonuclear cobalt bomb of up to 100 megatons against an enemy's naval ports and coastal cities.[46]

Military AI could provoke a new generation of asymmetric responses that will add threats to the world. There are virtually no legal restrictions on the use of naval drones in the world. Projects under development might have a significant impact on strategic stability and international security.

Proposals

Development of military AI obviously cannot be stopped. As it develops, international regulation is clearly needed to compel the military AI to follow human laws. It is time for states to consider new governance approaches to mitigate the possible risks of military uses of AI.

- It is necessary to define what experts mean by the term "military artificial intelligence." It is obvious that this is not a new type of weapons but rather a qualitative improvement to known types of weapons that gives them new capacities, like autonomy, better sensors, and reliable communication that ultimately allow the hardware to adapt to the environment.

- Where AI qualitatively improves known military systems, we might consider applying traditional arms control measures—transparency and confidence building.

- States could increase the transparency of global military activities by agreeing to publish data on new items in air, land, sea, and underwater arsenals with the function of unmanned operation, as it was specified in the Vienna Document for some conventional weapons. According to the Vienna Document, participants should exchange data on the main weapons and equipment systems in the zone of application for confidence- and security-building measures, as well as regular information on their plans to deploy them.

- In such transparency measures, states could agree to include data on demonstrations of new types of remote-controlled and autonomous weapons and prior notification of certain military activities.

In distant futures, military AI may gradually displace human responsibility in international security. Human societies may not traditionally compete with each other, but the whole of humanity may start competing with technology for the right to make decisions and determine its own destiny.

Artificial intelligence will gradually change the shape of states' militaries. Traditional military units will gain new strike capabilities through modern means of command and control, intelligence, accelerated collection and exchange of information, and automation of data processing. New service branches and new weapons systems have been introduced in a relatively short period of time, including missile defense systems, cybercommand, space forces, AI-based intelligence, surveillance and reconnaissance, information warfare, electronic warfare units, electronic countermeasures, laser weapons, autonomous vehicles, unmanned underwater vehicles, antidrone units, and hypersonic vehicles.

These war novelties serve as a signal that future conflicts will be more fleeting, lethal, sudden, and unpredictable. To date, humans have voluntarily delegated functions to artificial intelligence where it facilitates their work. Perhaps we are approaching a moment when decision making on defense and security will be increasingly delegated to artificial intelligence as a necessary measure, since limited human capabilities simply will not allow enough time for military and political leadership to make deliberate decisions. Before that future comes to pass, today is the time to make decisions on the safe future for our planet—as long as humanity is able to discuss it and compromise.

Endnotes

[1] Russia Today, "'Whoever Leads in AI Will Rule the World': Putin to Russian Children on Knowledge Day," September 1, 2017, https://www.rt.com/news/401731-ai-rule-world-putin/.

[2] Office of the President of the Russian Federation, "Decree of the President of the Russian Federation on the Development of Artificial Intelligence in the Russian Federation," Center for Security and Emerging Technology, translation, October 10, 2019, https://cset.georgetown.edu/wp-content/uploads/Decree-of-the-President-of-the-Russian-Federation-on-the-Development-of-Artificial-Intelligence-in-the-Russian-Federation-.pdf.

[3] Office of the President of the Russian Federation, "Meeting on the Development of Artificial Intelligence Technologies," May 30, 2019, http://en.kremlin.ru/events/president/news/60630.

[4] Ministry of Defense of the Russian Federation, "Scientific and Business Program of the International Military-Technical Forum," Army-2017, http://army2017.mil.ru/ndp.html.

[5] Ministry of Defense of the Russian Federation, "БОЕВОЙ РОБОТ (Battle Robot)," Russian Military Encyclopedic Dictionary, accessed October 2019, http://encyclopedia.mil.ru/encyclopedia/dictionary/details.htm?id=3551@morfDictionary.

[6] Igor Kalyaev and Ivan Rubtsov, "Боевым роботам нужна программа," *Национальная оборона*, 2012, http://www.oborona.ru/includes/periodics/defense/2012/0801/20258963/detail.shtml.

[7] A. Stepanov, "Отцы Ф.Ё.Д.О.Р.а: Минобороны безуспешно расходует деньги на создание боевых роботов," *Наша версия* 35 (September 13, 2018), https://versia.ru/minoborony-bezuspeshno-rasxoduet-dengi-na-sozdanie-boevyx-robotov.

[8] Office of the President of the Russian Federation, *Strategy of Scientific and Technological Development of the Russian Federations*, December 1, 2016, http://online.mai.ru/StrategySTD%20RF.pdf.

[9] A. Yudina interview with Sergey Anatolyevich, “Robotics Center of the Ministry of Defense of the Russian Federation: Microrobots of a ‘Pocket’ Format Will Appear in the Arctic,” Tass, August 24, 2017, https://tass.ru/interviews/4502372.

[10] Advanced Research Foundation, “About the Fund,” Advanced Research Foundation, accessed October 2019, https://fpi.gov.ru/about/.

[11] Ministry of Defense of the Russian Federation, “ERA Military Innovation Technopolis,” accessed October 2019, https://www.era-tehnopolis.ru.

[12] “Искусственный интеллект: проблемы и пути решения,” *Arsenal of the Fatherland* 33, no. 1 (March 2018), http://arsenal-otechestva.ru/article/1010-iskusstvennyj-intellekt-problemy-i-puti-resheniya.

[13] “Defense Ministry Will Combine Air Defense Divisions with EW Battalions,” *Военное обозрение*, August 22, 2018, https://en.topwar.ru/145917-minoborony-obedinit-divizii-pvo-s-batalonami-rjeb.html.

[14] Tass, “Kalashnikov Gunmaker Develops Combat Module Based on Artificial Intelligence,” July 5, 2017, https://tass.com/defense/954894.

[15] Kalashnikov Media, “«Калашников» представил высокоточный ударный беспилотный комплекс «ZALA ЛАНЦЕТ»,” June 24, 2019, https://kalashnikov.media/video/technology/kalashnikov-predstavil-vysokotochnyy-udarnyy-bespilotnyy-kompleks-zala-lantset.

[16] White House Office of the Press Secretary, “Soviet-United States Joint Statement on Future Negotiations on Nuclear and Space Arms and Further Enhancing Strategic Stability,” Janury 6, 1990, transcript accessed online via George H. W. Bush Presidential Library and Museum, https://bush41library.tamu.edu/archives/public-papers/1938.

[17] C. Hagel, “Reagan National Defense Forum Keynote,” November 15, 2014, https://dod.defense.gov/News/Speeches/Speech-View/Article/606635/.

[18] See Vincent Boulanin and Maaike Verbuggen, *Mapping the Development of Autonomy in Weapons Systems*, Stockholm International Peace Research Institute, November 2017, 39-41, https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117_1.pdf; Robert Work, “Ending Keynote: Art, Narrative, and the Third Offset,” remarks at the Global Strategy Forum: America’s Role in a Changing World, Atlantic Council, Washington, DC, May 3, 2016, https://www.youtube.com/watch?v=R9PswCdTi2E.

[19] Tass, “General Staff Warns US ABMs Can Detect Any Missile, Even Russian Ones,” March 28, 2017, https://tass.com/defense/937949.

[20] Zvezda, “Тени над морем: Россия ставит на вооружение уникальные противокорабельные ракеты,” December 5, 2014, https://tvzvezda.ru/news/forces/content/201412050825-6ctb.htm.

[21] L. Thompson, “A Stealthy Anti-Ship Missile Could Help U.S. Turn the Table on Chinese Navy,” *Forbes*, December 3, 2018, https://www.forbes.com/sites/lorenthompson/2018/12/03/lockheed-martin-stealthy-anti-ship-missile-fills-gap-in-u-s-navys-china-strategy/#5509884d51e0.

[22] Interfax, “Минобороны РФ признало ПРО США угрозой для низкоорбитальных спутников,” March 28, 2017, https://www.interfax.ru/world/555702.

[23] P. Tucker, “How AI Will Transform Anti-Submarine Warfare,” Defense One, July 1, 2019, https://www.defenseone.com/technology/2019/07/how-ai-will-transform-anti-submarine-warfare/158121/?oref=defenseone_today_nl.

[24] T. Rogoway, “Air Force’s Secretive XQ-58A Valkyrie Experimental Combat Drone Emerges After First Flight,” The Drive, March 6, 2019. https://www.thedrive.com/the-war-zone/26825/air-forces-secretive-xq-58a-valkyrie-experimental-combat-drone-emerges-after-first-flight.

[25] M. Davis, “The F-35: The Last Manned Fighter Aircraft?,” Fortuna’s Corner, August 8, 2016, https://fortunascorner.com/2016/08/09/the-f-35-the-last-manned-fighter-aircraft/.

[26] B. Chow, “Two Ways to Ward Off Killer Spacecraft,” Defense One, July 30, 2019. https://www.defenseone.com/ideas/2019/07/two-near-term-ways-ward-killer-spacecraft/158820/l.

[27] N. Guibert, “Comment la France va Militariser sa Doctrine dans l’Espace,” *Le Monde*, July 25, 2019, https://www.lemonde.fr/international/article/2019/07/25/la-france-militarise-sa-politique-spatiale_5493327_3210.html.

[28] Ministry of Defense of the Russian Federation, “Национальный центр управления обороной Российской Федерации,” https://structure.mil.ru/structure/ministry_of_defence/details.htm?id=11206@egOrganization; A. Leonkov, “Технические аспекты управления войсками России и США,” *Arsenal of the Fatherland* 21, no. 1, June 24, 2016, https://arsenal-otechestva.ru/article/753-tekhnicheskie-aspekty-upravleniya-vojskami-rossii-i-ssha.

[29] *Nezavisimaya Gazeta*, “Министерство обороны РФ форсирует переход на ‘цифру,’” May 25, 2019, http://nvo.ng.ru/realty/2019-05-24/2_1045_transition.html.

[30] V. Burenok, “Применение искусственного интеллекта в военном деле,” *Arsenal of the Fatherland* 33, no. 1, March 27, 2018, http://arsenal-otechestva.ru/article/1010-iskusstvennyj-intellekt-problemy-i-puti-resheniya.

[31] ЯRobot, “Итоги конференции ‘Искусственный интеллект: проблемы и пути решения’ 14-15 марта в парке ‘Патриот,’” March 22, 2008, https://ya-r.ru/2018/03/22/itogi-konferentsii-iskusstvennyj-intellekt-problemy-i-puti-resheniya-14-15-marta-v-parke-patriot/.

[32] I. Korotchenko, “Фонд перспективных исследований в структуре внешнего окружения,” January 4, 2013, https://vpk.name/news/87141_fond_perspektivnyih_issledovanii_v_strukture_vneshnego_okruzheniya.html.

[33] Office of the President of the Russian Federation, “Decree of the President of the Russian Federation About the Strategy of Scientific and Technological Development of the Russian Federation,” December 1, 2016, http://kremlin.ru/acts/bank/41449.

[34] “The Onrushing Wave,” *The Economist*, January 18, 2014, https://www.economist.com/briefing/2014/01/18/the-onrushing-wave.

[35] CBInsights, “The Race for AI: Here Are the Tech Giants Rushing to Snap Up Artificial Intelligence Startups,” September 17, 2019, https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/.

[36] M. Wheatley, “DARPA Plans to Spend Billions on AI Research for Weapons and Other Uses,” SiliconAngle, September 9, 2018, https://siliconangle.com/2018/09/09/darpa-says-plans-spend-billions-ai-research-weapons-uses.

[37] Z. Fryer-Biggs, “The Pentagon Plans to Spend $2 Billion to Put More Artificial Intelligence into Its Weaponry,” The Verge, September 8, 2018, https://www.theverge.com/2018/9/8/17833160/pentagon-darpa-artificial-intelligence-ai-investment.

[38] N. Strout, “Can Commercial Satellites Revolutionize Nuclear Command and Control?,” C4ISRNET, July 12, 2019, https://www.c4isrnet.com/battlefield-tech/c2-comms/2019/07/12/can-commercial-satellites-revolutionize-nuclear-command-and-control/.

[39] J. Corrigan, “Report: Pentagon Should Assume US Satellites Are Already Hacked,” Defense One, July 4, 2019, https://www.defenseone.com/technology/2019/07/report-pentagon-should-assume-us-satellites-are-already-hacked/158215/?oref=defenseone_today_nl.

[40] Nan Tian, Aude Fleurant, Alexandra Kuimova, Pieter D. Wezeman, and Siemon T. Wezeman, “Trends in World Military Expenditure, 2018,” Stockholm International Peace Research Institute, April 2019, https://www.sipri.org/sites/default/files/2019-04/fs_1904_milex_2018_0.pdf.

[41] Fabrizio Dragosei and Paolo Valentino, “Putin: ‘Ready to Talk to the US. In Constant Contact with Salvini’s League,’” July 4, 2019, https://www.corriere.it/esteri/19_luglio_04/putin-ready-to-talk-to-the-us-constant-contact-with-salvini-s-league-157f245e-9dec-11e9-9326-3d0a58e59695.shtml.

[42] Organization for Economic Cooperation and Development, “Research and Development Statistics Database,” 2020, https://stats.oecd.org/Index.aspx?DataSetCode=GBARD_NABS2007.

[43] Kirill Bulanov and Tatyana Lomskaya, “Эксперты назвали незаконной высокую долю секретных расходов бюджета,” Ведомости, October 17, 2018, https://www.vedomosti.ru/economics/articles/2018/10/17/783943-dolyu.

[44] S. Cao, “A.I. Apocalypse Yet? How World Leaders at Davos Have Changed Their Thoughts on Robots,” *Observer*, January 1, 2019, https://observer.com/2019/01/a-i-apocalypse-yet-how-world-leaders-at-davos-have-changed-their-thoughts-on-robots.

[45] Tass, “Hypersonic and Bottom-Based Missiles Make Part of Russian Troops’ Arsenal,” November 21, 2017, https://tass.com/defense/976672.

[46] V. Putin, “Presidential Address to Federal Assembly,” Gostiny Dvor, Moscow, February 20, 2019, http://en.kremlin.ru/events/president/news/59863; Tass, “Russia Floats Out First Nuclear Sub That Will Carry Poseidon Strategic Underwater Drones,” April 23, 2019, https://tass.com/defense/1055188.

About the Author

Vadim Kozyulin, PhD, is the Project Director of the Emerging Technologies and Global Security Project at the PIR Center (Russian Center for Policy Research) and a professor of the Academy of Military Science.

This working paper was prepared for a workshop, organized by the Stanley Center for Peace and Security, UNODA, and the Stimson Center, on The Militarization of Artificial Intelligence.