Assessing the Strategic Effects of Artificial Intelligence

Workshop Summary

September 2018

CGSR

Center for Global Security Research

LAWRENCE LIVERMORE NATIONAL LABORATORY

Workshop Summary

Assessing the Strategic Effects of Artificial Intelligence

Center for Global Security Research
Lawrence Livermore National Laboratory
&
Technology for Global Security

September 20-21, 2018

Paige Gasser, Rafael Loss, Andrew Reddie*

Workshop Concept

On September 20-21, the Center for Global Security Research (CGSR) at Lawrence Livermore National Laboratory (LLNL), in collaboration with Technology for Global Security (Tech4GS), hosted a workshop to examine the implications of advances in artificial intelligence (AI) on international security and strategic stability. Participating policymakers, scholars, technical experts, and representatives of various private sector organizations addressed the central question of whether the United States government should consider adjusting its approach to nuclear deterrence and strategic stability in light of the wide range of developments in the AI field. The workshop examined the potential risks and opportunities presented by military applications of AI and assessed which of these require consideration in the near term—and which might be exaggerated.

For the purposes of the workshop, we took a broad view of potential future applications of AI, including enablers of autonomous action; tools for decision support, simulation and modeling; and tools for collecting and analyzing very large volumes of information. We sought to understand the differences between near term impacts and potential longer-term possibilities, which are of course more difficult to forecast.

We began with an effort to establish a common understanding among participants about (1) the possible "strategic effects," including for example impacts on strategic stability and nuclear deterrence and (2) the main technological innovations that have come to comprise the field of "artificial intelligence." Second, participants compared and discussed the different approaches countries take in developing and adopting militarily relevant AI technologies. This part of the discussion focused on the United States, China, and Russia, but we recognize that AI can be expected to be developed and/or applied by many other countries. Then, the group examined

* *This report is a summary of workshop discussion. The views reported here are the personal views of individual participants and should not be attributed to the organizers, the National Nuclear Security Administration, the Department of Energy, or the United States government.*

AI's potential effects on cross-domain and multi-domain deterrence. Another panel explored whether battlefield AI and tactical advantages gained through AI-enabled technologies could produce strategic effects. Finally, participants assessed, based on the previous discussions, what the United States' approach should be to ensure strategic stability into the future.

Panel 1: Revisiting Strategic Stability and Recent Developments in Artificial Intelligence

- How do we, our allies, and adversaries define strategic stability in a time of renewed great power competition?
- What aspects of AI are most relevant for strategic stability? Will they strengthen or undermine deterrence?
- How rapidly should we expect AI-related technologies to proliferate? Will certain technologies spread more evenly than others?
- How might AI interact with other technological innovations, such as advances in quantum computing, in affecting deterrence dynamics?

When trying to assess the strategic effects of AI, direct evidence is in short supply. This task is complicated by the fact that there is considerable debate about the definition and requirements of strategic stability and about its continued utility as a concept for organizing U.S. interests and U.S. policy. In the Cold War, thinking focused on the risks of general war, of nuclear war, and of preparations for war that might make it more likely or more prone to escalate. The strategic relationship was defined as stable if there was no pressure for either side in a conflict to strike preemptively by nuclear means, and if neither side perceived the other as likely to gain a first strike advantage through force modernization. These concepts were developed in the context of bipolar confrontation under the shadow of nuclear Armageddon amidst decades of arms racing and after the lessons learned from the Cuban missile crisis. How AI assisted weapons fit in a world that is more multipolar than bipolar, where the risks of limited war appear to be on the rise, and major powers are reasserting nuclear postures is quite uncertain. What lessons from the Cold War translate into the new environment? The more multi-dimensional nature of strategic conflict has only magnified this debate, as competition for strategic advantage in cyber space and outer space have greatly complicated the calculus of benefit, cost, and risk associated with different courses of action.

The emerging AI competition among the major powers fits the old framework in two primary ways. First, the emergence today of that competition introduces a significant new element of uncertainty in the strategic calculus. Each side worries about the advantages that the other(s) may be gaining with AI development, so each increasingly feels compelled to compete, thus reinforcing the fear of others that some decisive new military advantage may be gained. This is analogous to the arms race instabilities of the Cold War.

Then there is the longer-term problem, where real advantage may be gained and exploited in crisis to put an enemy in an escalate-or-lose situation. This is analogous to the crisis instabilities of the Cold War. How realistic a concern is this? Future developments in AI technologies have the potential to significantly influence the military balance between strategic competitors. AI is poised to affect the nuclear and the conventional military balance in conflict zones by influencing the speed and information available to each side. One particular advantage likely to affect regional stability is the development of AI-guided Intelligence, Surveillance and Reconnaissance (ISR) systems that provide real-time information and analysis of the battlefield, vastly improving the targeting and lethality of autonomous systems operating on land, sea, air, space, and cyber and altering patterns of conflict escalation.

The global proliferation of AI and other dual-use technologies ensures widespread applications for military purposes. Growing investments in AI and its applications in technologies, such as robotics, autonomous vehicles, supercomputing, cyber warfare, and quantum computing all have military implications. China and Russia are making AI a priority and openly tout their successes. This year, China announced that it is spending $300 billion on AI technologies and Russian President Putin proclaimed that whoever succeeds in AI "will become the ruler of the world." Other countries are on track to make rapid progress over the next decade. Indeed, great power competition may be accelerating the adoption of AI for military purposes, causing a global crisis of conscience for engineers and researchers who reject contributing to the militarization of AI. Like it or not, some participants remarked, the AI arms race has already begun—an arms race that is dramatically different from a traditional nuclear arms race.

In contrast, other participants suggested that the actual international security consequences of AI have been exaggerated. The technology, they argued, is not ready for "prime time," and will not bring the hoped-for benefits advocated by proponents—at least not in the near-term. AI is plagued by a variety of daunting problems, including the unreliability of data, the propensity of AI neural networks to make mistakes, and the small performance window for algorithms designed to address specific challenges. Utilizing AI for lethal applications, they warned, is premature and risks a wide range of unpredictable outcomes. AI algorithms regularly misinterpret data and produce perplexing results. The disaggregation of data represents another technical consideration as there is a need for a stable mechanism to communicate and validate that data was accurately aggregated. More experimentation is necessary to understand how and why AI works and to gain confidence in its output. The inherent "black box" problem also prevents researchers from understanding how AI comes to conclusions based on the data it is given. This opacity-transparency problem means that it is difficult for operators to ascertain a distinguishable cause and effect relationship between the input data, the algorithms, and the outputs—a significant challenge when troubleshooting AI technology.

These characteristics have led to AI technologies being kept separate from essential security-related technological systems—particularly those designed to have kinetic effects (at least in the United States). Other countries may not be so restrained, and also may be more willing to mobilize their data unencumbered by norms regarding surveillance and privacy. Neither Russia nor China shares U.S. perspectives about the line between commercial and governmental applications as their companies gear up to support national agendas.

This gap between possible military applications of AI and concerns about its readiness for those applications was a consistent theme throughout the workshop. While acknowledging some countries are beginning to deploy AI informed systems already, several experts argued that AI requires more experimentation and development before being integrated into lethal weapon systems and their decision-making support systems.

Panel 2: Comparing AI Adoption and Integration Across Countries

- Does AI empower different actors in different ways? Do different actors follow different approaches in developing and adopting AI technologies?
- Does AI increase the potential for asymmetrical conflict and strategies?
- Are non-state actors leveraging AI technologies likely to serve as spoilers in this new strategic environment?
- What drives cooperation and/or competition among various actors? How could incentives be altered to enhance strategic stability?
- Is there an AI arms race? What does it look like and what does it take to win it?

In this panel, participants examined the development of artificial intelligence technologies in Russia and China to reflect upon an era of renewed strategic competition among the United States, Russia, and China.

With regards to Russia, participants pointed to President Putin's September 2017 statement noting that Moscow intends to marshal the intellectual and technical capital necessary to arrive at cutting-edge military technology—with the long-term goal of controlling the information space. These developments have been reflected in the creation of a series of Ministry of Defense (MoD)-affiliated research centers and conglomerates that have substantially increased Russia's national research base. Most notably, the Foundation for Advanced Studies (FAS) is playing an increasingly important role in shaping Russian AI development with image and speech recognition, control of autonomous weapons systems, supercomputing, robotics, and the analysis of weapon system lifecycles. The Russian private sector has also benefitted from government support in terms of human capital development and early investment in cutting-edge information technologies as it attempts to substitute imports with indigenous technologies—this is in spite of Russia missing a "start-up culture" and its continued dependence upon Western technology.

Participants also noted Moscow's comparative advantage regarding data collection and its ability to distort or poison its adversaries' AI training data—a form of information warfare. Priority AI applications include unmanned ground vehicles (UGVs), aerial unmanned systems, electronic warfare systems, various swarm technologies, and unmanned underwater vehicles (UUVs), some of which have been tested by Moscow in Syria and as part of the Vostok 2018 exercises.

With respect to China, too, participants noted the enormous investments that Beijing has been providing to technology firms as part of a broader effort to combat the U.S. third offset strategy. China's strategy, described as "offsetting the offset," involves developing quantum computing, advanced Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR), hypersonic capabilities, and AI technologies. These developments were described as part of a larger strategy by the PLA to catch-up and surpass the United States: "Whatever the enemy fears is what we must develop." This has led to the advancement of a hybrid model of innovation by mobilizing government investment and scaling human capital in support of the "national team" and "national champions." This has most recently been exemplified by the Made in China 2025 plan—one of many science and technology megaprojects supported by the Chinese state. Regarding AI specifically, the AI Development

Plan 2017 has led to a remarkably rapid innovation in both the field of artificial intelligence and ancillary disciplines, such as cognitive science and quantum computing.

Beijing's primary obstacle to AI development, however, is the availability of talent enabling a major national initiative. And while the notion that "China can't innovate" is no longer true, Beijing has attempted to use technology transfer schemes, partnerships with multinational companies, and repatriation of Chinese nationals educated abroad to bolster its capacity.

In both states, leaders are pursuing advanced technologies as part of their broader national security strategies. At the same time, perceptions and misperceptions related to competitor advancements vis-à-vis AI technologies in Washington, Moscow, and Beijing are likely to be central to future arms racing and crisis stability—with consequences for broader strategic stability.

Panel 3: Artificial Intelligence and Deterrence Across Domains

- How might AI affect the key components of strategic deterrence, such as C4ISR, the weapons complex, second strike capabilities, and space-based systems?
- How might AI technologies impact deterrence strategies across domains? What is the relationship between AI and integrated/complex deterrence?
- Do developments in AI technologies shift thinking about critical national security infrastructure? Do they shift the requirements for engagement between the private and public sectors?

This discussion reflected significant anxiety about the potentially damaging effects of AI-focused competition on the stability of nuclear deterrence and strategic deterrence more broadly. Stable mutual deterrence required opposing nuclear powers to have credible, survivable nuclear forces. Technological advances and geopolitical developments, however, have broadened the scope of the deterrence discourse with more of an emphasis on regional stability, conventional deterrence, and multi-domain warfare. Cyber and space-based capabilities have become critical to how the United States and other countries deploy and operate their conventional and nuclear forces. This has provided military and political leaders with new opportunities, but it has also revealed new vulnerabilities in systems and deterrence strategies. Accordingly, deterrence stability is increasingly dependent on an expanding multitude of conventional, allied, and regional capabilities. Complex, cross-domain, and multi-domain deterrence describe this growing connectivity between the ability to deter regional conflicts, escalation at the conventional level, and the credibility of strategic nuclear deterrence as space and cyberspace become more and more critical for conventional and nuclear systems.

These conceptual frameworks seek to leverage deterrent capabilities in one domain to deter adversarial activity in other domains. The increasingly complex nature of deterrence also affects the strategic stability calculus. AI is becoming more relevant for each component of the complex deterrence architecture, especially given its potential to significantly affect ISR and compress decision-making timelines, as well as introducing vulnerabilities mentioned in previous panels.

AI could have a significant effect on key areas of military innovation and force posture. For example, an increased ability to target strategic assets with conventional precision strike capabilities, enhanced through AI, might heighten instability, especially in crisis situations. Workshop participants noted that AI-enabled advances in ISR might provide technologically advanced powers with the ability, or at least perceived ability, to conduct disarming first strikes without using their own nuclear weapons. Improvements in ballistic missile defenses, also made possible with AI, might then negate remaining retaliatory efforts. Under these conditions, adversaries would be well-advised to put their nuclear forces on hair-trigger, "launch on warning" alert, invest in robust defensive systems, and double down on mobile launch systems that complicate detection in order to maintain a second-strike capable nuclear force.

It is highly unlikely, however, that launch authority over nuclear weapons would be knowingly delegated to AI decision making mechanisms. Leaders in all types of political systems - democratic and autocratic – would be inclined to maintain central authority over nuclear weapons and would not willingly delegate such momentous authority to automated systems. Whether such automated systems will always respond to human preferences is the subject of books and movies such as 2001 A Space Odyssey, Dr. Strangelove, and WarGames, but beyond the scope of the workshop.

AI-enabled technologies are well-suited to military applications beyond ISR and conventional strike. AI is already being used to improve logistics across the armed forces and is key to many simulation and modeling applications for the design and optimization of weapon systems and integration software. In cyber warfare, AI appears to be improving the speed and effectiveness of efforts to characterize and penetrate adversary networks. On the other hand, AI could help cyber defenders monitor intrusions and detect anomalies, such as malware and implants in networks and operating systems. AI-enabled signals processing could improve the performance of early warning radar and be used for situational awareness in space. These improvements in cybersecurity to protect critical command, control, and communication systems could potentially strengthen regional and strategic stability.

Participants also discussed possible applications of AI to aid in arms control verification. They were skeptical, however, that the spread and use of potentially destabilizing AI technologies could be controlled through traditional arms control mechanisms. The dual-use nature of most AI would make verification nearly impossible.

While none of the near-term applications of AI to multi-domain deterrence represent fundamental changes in the logic of deterrence, some technologies have the potential to inspire countermeasures that could make crisis management more difficult. The speed at which AI-guided ISR platforms could direct and execute kinetic operations could limit options for de-escalation. Workshop participants agreed that these issues are ripe for discussion at future strategic stability talks between the United States, Russia, and China.

Panel 4: Operationalizing Automation and Artificial Intelligence for the Battlefield

- How might AI change the character of conflict, its initiation, escalation and termination?
- How might it affect the offense-defense balance in major power and regional conflict as well as in campaigns involving non-state actors?
- Could tactical battlefield AI aggregate up and have strategic effects?
- How might non-state actors leverage AI technologies to threaten state actors? What are the counter-AI/adversarial AI tools needed to mitigate these risks?

In this portion of the discussion, expectations were high, but the caveats were numerous. This panel began with the observation that integrated systems that track operational data are not a recent invention—recall the demand in WWII for smart munitions and widespread use of data analytics for a variety of traditional military purposes, such as logistics and supply chain management. While agreeing that AI may have tremendous implications on the battlefield (i.e. changing the character of conflict and escalation), some participants urged caution in thinking about and preparing to use AI in future warfare. Technology experts warned that AI is still largely untested and not reliable enough to be released "into the wild." Data collections are fragile and easily polluted with incorrect information. Moreover, AI has also shown a propensity to both reflect human bias and create surprising distortions on its own.

The tolerance for such unpredictable results differs with the application. Military applications will find them especially problematic. Unproven AI technologies could lead to unexpected and undesirable outcomes on the battlefield. On the other hand, AI is already proving useful for some missions, such as object recognition, simulation, and analysis, but is not ready for widespread use in lethal platforms. Additional research and experimentation remain necessary before AI can be trusted on the battlefield.

Despite these concerns, there is little doubt that AI-guided autonomous vehicles are poised to change the face of warfare. Highly-integrated ISR coupled with autonomous systems operating in multiple domains will accelerate the pace of battle and give tactical advantages to whoever can collect, process, analyze, and operationalize data faster than their adversary. The speed and precision gained with AI, however, will be constrained by the need to keep a "human in the loop." This leads to a number of important questions: Where in the kill chain should the human be injected, and with what authorities? How might human decision making detract from the advantages gained from AI and autonomy? Recounting the establishment of DoD's "Project Maven"—a program which focuses on computer vision to autonomously extracts objects of interest from moving or still imagery—participants warned against removing the human from the loop in the service of enhanced warfighting speed.

Panelists noted potential negative impacts of conducting kinetic operations at the accelerated pace made possible by AI and autonomy. They also noted potential ethical considerations, which are already relevant in connection with the growing use of drones. The lack of norms raises troubling issues regarding the future rules of warfare. Another consideration is the risk of escalation. Surprise attacks and blitzkrieg offense, if not immediately decisive, are likely to raise the stakes of war and provoke escalated—and potentially automatic—responses. Escalation at the tactical level could easily become strategic.

The conversation then shifted to how potential adversaries are thinking about battlefield applications of AI. Adversarial AI, where counter-AI methods are used to assess and exploit vulnerabilities in AI systems, could greatly complicate AI usage on the battlefield. For example, specific "blind spots" in neural networks could allow an adversary to nefariously tweak data in such a manner that a human and a neural net would not recognize a change—this phenomenon is described as "data pollution." Many countries are investing in AI, recognizing they cannot afford to miss out on potentially game-changing military applications. Different national standards will likely result in a broad range of AI methods being deployed, some of which could be even less reliable than those currently being contemplated by the United States and other major powers.

All of the problems discussed in this and previous panels could be magnified when an increasing number of countries with diverse approaches to command and control and the "human in the loop" issue use unreliable and unpredictable AI technologies. Counter-AI measures might undermine many aspects of AI-supported military operations, possibly leading to mistakes in targeting. Hitting the wrong targets, including mistaking civilian infrastructure for military installations, could hasten escalation and broaden conflicts. The mistaken destruction of the Chinese embassy in Belgrade—though not caused by an AI system—offers an example of accidents having broader and long-term political implications. Securing data against hacking is already a global problem. Protecting dynamically changing battlefield data from counter-AI actions and imposing costs so potential adversaries do not use AI as a tool to exploit the United States presents daunting challenges. Many of those challenges, including adversarial AI, remain unsolved even by the best technical AI minds in the world.

Finally, the group discussed ways that AI-fueled battlefield operations could promote or hinder regional stability, particularly in Asia and Europe. To the extent that AI reinforces the credibility of conventional deterrence, stability results. However, unpredictability and mistakes in a regional conflict could increase risks of escalation and undermine stability.

Panel 5: Ensuring Strategic Stability in the Age of Artificial Intelligence

- What are U.S. priorities for taking advantage of AI to enhance deterrence?
- What are the main risks that AI poses for strategic stability?
- How might cooperation help steer AI to enhance, rather than undermine, national and international security?

Having completed the various "deep dives" on panels 2-4, the final panel returned to the opening questions about strategic stability and the additional questions associated with actions that the United States can and should take.

In addition to issues of conflict escalation, participants noted the potential for AI to enable a credible first strike capability and degrade an adversary's second-strike capability. The ability to locate and hit an adversary's strategic assets could disrupt the fundamental tenets of deterrence and erode the foundations of mutual assured destruction by shifting incentives away from accepting mutual vulnerability to exploiting vulnerability and striking first to eliminate retaliatory forces. As discussed in earlier panels, the actual ability to successfully conduct disarming first strikes is premature. However, the perception that an adversary might even consider such a strike could be destabilizing with the potential to lead to miscalculation and likely inspire counter-measures. Participants cited Vladimir Putin's boasting about new Russian nuclear forces capable of evading defenses as an effort to maintain assured retaliation. Another possibility would be for weak nations fearing advanced AI informed weapons to resort to asymmetric means in an effort to "level the playing field."

Some participants raised the issue of AI's role in information warfare and influence operations and its possible implications for strategic stability. AI-supported bots, fake news, and social media campaigns could distort public perception and interfere with critical communications in times of conflict. These efforts could have destabilizing effects, especially in countries where public opinion translates into political power. Political leaders may be influenced by public sentiments and make decisions about war and peace based on fictitious or fabricated narratives. AI cyber operations also have the potential to shape decisions to use force, make threats, or escalate conflict. Beyond the U.S.-Russia and U.S.-China deterrence dyads, other nuclear armed countries, such as India and Pakistan, should not be overlooked as they would be subject to the same types of manipulation and escalation dynamics. Deterrence, participants noted, depends on clear and effective communication.

Ultimately, the goals of strategic competition in AI remain obscure. AI magnifies existing concerns about deterrence stability, escalation management, and conflict resolution. Established methods developed to support these objectives, such as arms control and hotlines between leaders, may not be easily adaptable to the risks that appear to be aggravated by AI and appear to be ill-suited to manage these new challenges to instability. Moreover, since AI is not a siloed or monolithic technology, policymakers cannot reflect on the implications of AI without considering big data, high performance computing, the proliferation of sensors in society and on the battlefield, as well as advancements in robotics and quantum computing.

AI can be viewed as part of a new information-oriented revolution in military affairs. It appears poised to affect a wide range of military technologies, thereby changing the way future wars will be fought and won. While an AI arms race has begun, AI may be best viewed as an enabling technology rather than a disruptive technology in its own right. The ability to process and draw meaning from massive data sets is evolving rapidly, but effectively operationalizing these evolutionary developments must consider various organizational, economic, cultural, physical, and psychological realities. These factors are especially relevant for military applications where the stakes for national security and strategic stability are high.

The vital role of the private sector in the development of AI technologies brings both opportunities and challenges from a national security perspective. Reliance on the private sector for AI technology and expertise will influence how it is applied. Leading technology companies are eager to conduct business with governments around the world. Countries, such as China, that exert more control over their science and technology sectors may have advantages in adapting AI for military purposes. Russia's deficiencies in private sector technology investment presents challenges. In the absence of controls on the proliferation of AI, competition will be tenacious.

The closing note was one of anticipation. The diverse workshop participants—policymakers, scholars, technical experts, and representatives of various private sector organizations—shared a common view that we are standing at the beginning of a long journey, as we attempt to understand the practical, moral, legal, and public policy implications of AI and as we attempt to shape those factors for the common good. Although it is too early to make definitive judgments about the long-term effects of AI on strategic stability, the workshop identified several major factors that will shape how AI influences the military strategies that form the basis of deterrence.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-TR-761425.