POLICY ROUNDTABLE:

Artificial Intelligence and International Security

June 2, 2020

Table of Contents

1. "Introduction: Artificial Intelligence and International Security," By Michael C. Horowitz, Lauren Kahn, and Christian Ruhl
2. "The AI that Wasn't There: Global Order and the (Mis)Perception of Powerful AI," By Mary (Missy) Cummings
3. "Integrating Emerging Technology in Multinational Military Operations: The Case of Artificial Intelligence," By Erik Lin-Greenberg
4. "The Militarization of Artificial Intelligence," By Paul Scharre
5. "The Promise and Risks of Artificial Intelligence: A Brief History," By Rebecca Slayton

1. Introduction: Artificial Intelligence and International Security

Michael C. Horowitz, Lauren Kahn, and Christian Ruhl

Advances in artificial intelligence (AI) have the potential to shape the global order, from the character of war to the future of work.[1] In response, countries and institutions are planning for this new AI world.[2] To assess the effects AI is having on the global order, Perry World House, the University of Pennsylvania's hub for global affairs, convened a two-day colloquium last autumn titled "How Emerging Technologies Are Rewiring the Global Order." As part of the colloquium, academics, policymakers, industry leaders, and other experts assessed how emerging technologies like AI are changing international politics. One panel focused on AI and international security, tackling questions such as "How well do militaries and/or the international security community understand the military and non-military effects of AI on international security?" and "How are advances in AI likely to shift the trajectory of great power competition?"

The articles in this roundtable represent analyses that colloquium panelists Missy Cummings, Erik Lin-Greenberg, Paul Scharre, and Rebecca Slayton wrote in response to these questions prior to the colloquium.

[1] Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 1, no. 3 (May 2018): 36–57, https://doi.org/10.15781/T2639KP49.

[2] "National and International AI Strategies," Future of Life Institute, accessed Dec. 23, 2019, https://futureoflife.org/national-international-ai-strategies/.

What Is AI?

The U.S. Department of Defense defines AI as "the ability of machines to perform tasks that normally require human intelligence," such as "recognizing patterns, learning from experience, drawing conclusions, [and] making predictions."[3] But to what extent does discussing the effects of AI on the international security environment require consensus on what constitutes AI in the first place? While most agree that AI is a general-purpose enabling technology with the potential to shape global politics, the nuances of what developments, technologies, and historical precedents actually fit under the AI umbrella vary greatly, even within the papers presented here: While Slayton cites AI methods and approaches as going back nearly seven decades, Scharre dates AI as relatively new.

To overcome some of the definitional and taxonomical difficulties, Scharre suggests utilizing three different lenses through which to examine AI: specific applications, historical analogies, and technical characteristics. The contributors to this roundtable view AI through all three lenses and expose several risks and challenges of AI as it relates to international security.

AI Risks and Challenges

Some of the biggest ways in which AI could influence international peace and security stem not from superintelligence or "killer robots," but from the risks inherent in trusting computer-generated algorithms with making choices that humans used to exclusively

[3] *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, U.S. Department of Defense, last updated Feb. 12, 2019, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF

make. For example, as Scharre explains in his essay, future research should focus on specific applications of AI because AI is not a weapon or weapons system but is instead a technology that can enhance other technologies. There are, however, multiple risks and challenges common to a wide range of AI applications. Here we outline several such issues that are relevant to international security, including data risks, brittleness, integration, and the potential for misperception and misunderstanding.

Data Risks: Sharing, Biases, and Poisoning

Prominent risks in AI result from machine learning's requirement of large amounts of human-generated data for training. As Lin-Greenberg notes, data sharing faces political and technical issues. Politically, states may be unwilling to share sensitive data even with allies, for fear of accidentally disclosing too much information, especially on security issues.

Even when data are available, however, they often inadvertently include systemic racial, gender, and other biases. Computer vision technologies, including those from prominent companies such as Google, have infamously misidentified black people as gorillas or completely failed to recognize non-white individuals.[4] As colloquium participants pointed out, algorithmic biases that are more difficult to identify may be just as insidious. Machine learning algorithms designed to aid in criminal risk assessments, for instance, may learn racial bias from historical data, which reflects racial biases in the American criminal justice system. One study of a commonly used tool to identify criminal

[4] Kate Crawford, "Artificial Intelligence's White Guy Problem," *New York Times*, June 25, 2016, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?ref=technology.

recidivism found that the algorithm was 45 percent more likely to give higher risk scores to black than to white defendants.[5] Militaries that are using AI-enabled technologies must be wary of the tendency to reinforce the biases of the data they were trained on.

Systemic biases in datasets present a major problem for AI applications in international security, as they could prevent military applications from operating effectively and safely. Bad data, however, is not always accidental. "Data poisoning" attacks would allow adversaries to manipulate algorithms by injecting bad data into training datasets. Programs like the Intelligence Advanced Research Projects Activity's TrojAI and SAILS seek to address this problem: TrojAI attempts to build programs that can identify when algorithms have been trained on poisoned data, while SAILS is intended to keep attackers from finding sensitive training data in the first place. However, the threat remains widespread, especially as algorithms become more complex and their training data more voluminous.[6] Data poisoning attacks could also exacerbate other risks in AI, like its brittleness or inability to handle new and uncertain environments.

Brittleness and Performance in Uncertainty

Narrow AI systems — those artificial intelligence systems that are specified to handle a singular or limited function — can learn specific tasks, and learn them well. Game-playing algorithms, for example, have been outperforming humans at increasingly complicated games since the 1950s, as Slayton notes. Cummings explains that machines and

[5] Jeff Larson, Surya Mattu, and Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm," *ProPublica*, May 23, 2016, https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

[6] Jack Corrigan, "The Pentagon Wants to Stop Enemies from 'Poisoning' AI," *Nextgov*, last modified Jan. 25, 2019, https://www.nextgov.com/emerging-tech/2019/01/pentagon-wants-stop-enemies-poisoning-ai/154431/.

computers excel in controlled and simulated environments, such as in the case of AlphaGo and IBM Watson's *Jeopardy!* debut. Machines are far better than humans at searching a large set of known options and adhering to deterministic parameters.[7] These widely publicized successes represent undeniable progress in AI and machine learning, but they apply only in the carefully controlled environment of the game or specific narrow task.

AI algorithms themselves are often brittle — powerful, but liable to shatter when operated outside of a deterministic domain. Paul Scharre and Michael Horowitz argue that current AI systems "lack the general-purpose reasoning that humans use to flexibly perform a range of tasks," and can be overwhelmed and fail if they are "deployed outside of the context for which they were designed."[8] In the case of AlphaZero, Scharre and Horowitz warn that different versions of the algorithm needed to be trained for each game, and it "could not transfer learning from one game to another, as a human might."[9] Even more extreme, Scharre notes in his contribution to this roundtable that despite the ability of AI systems to "achieve superhuman performance in some settings," not only can they not transfer learning, but they can "fail catastrophically in others." In military settings, this can be particularly devastating. Functioning outside of a lab requires, as Cummings notes, "drawing abstract conclusions that require judgment under conditions of uncertainty." In national security applications, almost all situations are clouded by the fog of war, generating uncertainty that makes the effective use of AI more challenging.

---

[7] John Markoff, "Computer Wins on 'Jeopardy!': Trivial, It's Not," *New York Times*, Feb. 16, 2011, https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html.

[8] Paul Scharre and Michael Horowitz, "Artificial Intelligence: What Every Policymaker Needs to Know," Center for a New American Security, June 19, 2018, https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know.

[9] Scharre and Horowitz, "Artificial Intelligence: What Every Policymaker Needs to Know."

The fast-changing and unpredictable nature of conflict may be incompatible with brittle AI-enabled systems that cannot adapt the way a trained human soldier could. Similar to self-driving cars, Cummings argues that advanced military applications of AI require "a safety driver behind the wheel."

Many military applications of AI are still more theory than reality.[10] As Cummings discusses in her contribution, "There are very few actively deployed military systems that rely on AI" and those that do exist, like the primitive AI in the Tomahawk missile system, struggle with uncertainty and scene changes. Even the new bomb designed by Israeli Rafael Advanced Defense Systems (publicly disclosed in June 2019), which boasts an AI and Deep-Learning Advanced Target Recognition feature, really only uses AI in the final stages of guidance in order to home in on the predefined target.[11]

Integration and Compatibility

Success in the adoption of AI systems in military contexts will require significant systems integration and familiarity by human operators. During the colloquium, most panelists agreed that "some of AI's greatest implications in the military realm will be how command and control structures will need to adapt to integrate such cognitive capacity

[10] "Autonomous Weapons: An Open Letter from AI & Robotics Researchers," Future of Life Institute, July 28, 2015, https://futureoflife.org/open-letter-autonomous-weapons/.

[11] "Rafael Unveils New Artificial Intelligence and Deep-Learning Technologies in SPICE-250 to Enable Automatic Target Recognition," Rafael Advanced Defense Systems LTD., June 10, 2019, https://www.rafael.co.il/press/elementor-4174/.

into decision-making processes."[12] There are three levels at which a lack of integration and socialization of AI can cause breakdowns: 1) between new AI systems and legacy systems; 2) between the human operators and decision-makers and AI systems; and 3) between organizations that utilize AI systems.

Achieving cohesion at levels one and two will challenge individual organizations. They will have to adapt current command-and-control processes to best leverage AI technology and to hedge against brittleness or failures due to miscommunication between humans or systems. As Lin-Greenberg notes, however, interoperability may suffer due to the differences in the rate of diffusion of AI technology across countries, making it more challenging for a country to work with its partners and allies. Integrating AI technologies into these existing systems and dynamics will require significant organizational and technical changes, as well as flexibility and patience on the part of the organizations. Military leaders and decision-makers must not rush these processes.

Misperception and Decision-making

A fourth set of risks relates to misperceptions *about* AI's capabilities and misperceptions *as a result of* AI-enabled decision-making. All of the contributors to this roundtable highlight the problem of overconfidence. Overconfidence in the possibilities of AI, for instance, may lead militaries to apply machine learning to situations that are too complex for brittle algorithms. In human-machine teams, moreover, overconfidence in a computerized system can lead to "automation bias," where humans cognitively offload

[12] Christine Carpenter, Casey Mahoney, and Christian Ruhl, *Fall 2019 Colloquium Report: How Emerging Technologies Are Rewiring the Global Order*, Perry World House, 2019, 14, https://drive.google.com/file/d/1vK72tS3rA89DopZ_uE6DIyilD41n8kDX/view.

judgment to algorithms that they do not realize have flaws.[13] Such automation bias may lead human operators to miss the machine's false negatives and false positives because the operators are overconfident in the machine's ability.

Misperception also extends to a state's assessment of its adversaries' supposed capabilities. As Cummings' contribution to this roundtable shows, AI capabilities are difficult to verify. States could easily copy Silicon Valley's "fake it till you make it" strategy, in which human operators create the illusion of superior machine intelligence.[14] Possibly inflated claims will raise large challenges for intelligence agencies, and, from an analysis perspective, warrant vigilance and skepticism.

Strategy and Policy Recommendations

Nearly 88 percent of colloquium participants believed that an arms race over military applications of AI within the next 15 years is somewhat or very likely. The language of an arms race is potentially inappropriate in this context, however, even if popular and evocative.[15] Instead, it makes more sense to discuss how AI could impact international competition. Nevertheless, if countries believe they are in an AI arms race, this could generate risks if they race to the most advanced AI capabilities without making

[13] Michael C. Horowitz, Paul Scharre, and Alexander Velez-Green, "A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence," arXiv, December 2019, https://arxiv.org/pdf/1912.05291.pdf.

[14] Olivia Solon, "The Rise of 'Pseudo-AI,'" *The Guardian*, July 6, 2018, https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies.

[15] Heather M. Roff, "The Frame Problem: The AI 'Arms Race' Isn't One," *Bulletin of the Atomic Scientists*, April 29, 2019, https://thebulletin.org/2019/04/the-frame-problem-the-ai-arms-race-isnt-one/. After all, we would not speak of an "electricity arms race," to take a general-purpose enabling technology that AI is often compared to as an example.

appropriate and corresponding investments in safety, transparency, regulation, and ethics. If countries prioritize speed, they risk launching potentially more unreliable and more volatile systems than a measured pace might yield, as well as rushing the integration processes required to properly leverage the technology. In understanding possible arms-race-like dynamics, therefore, perception again matters as much as reality. As Paul Scharre has written, "the perception of a race will prompt everyone to rush to deploy unsafe AI systems" that do not adequately address the risks explored in this roundtable.[16] In other words, the AI arms race could wind up becoming a "race to the bottom." Furthermore, Lin-Greenberg's essay sheds light on how too much focus on an "AI arms race," and more broadly the focus on leading AI states like China and the United States, may crowd out important conversations on how advances in AI may leave allies behind, exacerbate global inequality, and widen the digital divide.[17]

No matter whether one believes there is an AI arms race or not, national investments in AI are clearly growing. The U.S. budget for Fiscal Year 2020, for instance, singled out AI as a research and development priority and proposed $850 million of funding for the American AI Initiative, spread out over several agencies.[18] Investments by the Chinese

[16] Paul Scharre, "The Real Dangers of an AI Arms Race," *Foreign Affairs* 98, no. 3 (May/June 2019), https://www.foreignaffairs.com/articles/2019-04-16/killer-apps.

[17] Mike Butcher, "World Economic Forum Warns of AI's Potential to Worsen Economic Inequality," *TechCrunch*, Jan. 23, 2019, https://techcrunch.com/2019/01/23/world-economic-forum-warns-of-ais-potential-to-worsen-global-inequality/.

[18] "Research and Development," The White House, accessed Dec. 23, 2019, https://www.whitehouse.gov/wp-content/uploads/2019/03/ap_21_research-fy2020.pdf.

government, which has declared its intentions to lead the world in AI by 2030, are estimated to exceed tens of billions of dollars.[19]

As Cummings argues below, states should "arm themselves" not just with greater investments in innovation and military applications, but with the capabilities to understand the weaknesses of AI-enabled warfare and spot inflated claims of superiority. As Slayton notes, the current Defense Department AI strategy focuses on "a culture of experimentation and calculated risk taking," but using AI effectively requires not just innovation, but also "continual maintenance of intelligent systems to ensure that the models used to create machine intelligence are not out of date." Innovation without maintenance and safety may increase both vulnerabilities and accidents.

Safety, after all, is another way of saying effectiveness, and militaries more than anyone want their weapons systems to work. Brittle or biased AI is not only prone to accidents, but vulnerable to deception. One research team at MIT demonstrated how a model of a turtle can be altered to reliably fool a computer vision algorithm into thinking it's a rifle.[20] To defend against such spoofing and to make machine learning applications safer, states should invest more resources in programs like the Defense Advanced Research Projects Agency's Guaranteeing AI Robustness against Deception (GARD).[21] GARD seeks to create AI that is more resistant to deception by funding research that builds theoretical

[19] Gregory C. Allen, "Understanding China's AI Strategy," Center for a New American Security, last modified Feb. 6, 2019, https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy.

[20] Will Knight, "Military Artificial Intelligence Can Be Easily and Dangerously Fooled," *Wired*, Oct. 21, 2019, https://www.technologyreview.com/s/614497/military-artificial-intelligence-can-be-easily-and-dangerously-fooled/.

[21] "Defending Against Adversarial Artificial Intelligence," DARPA, last modified Feb. 6, 2019, https://www.darpa.mil/news-events/2019-02-06.

foundations for defensible AI, actually creates these systems, and produces "testbeds" for evaluating them. The GARD program will start with images and progress to more complex problems of video and audio.[22]

Hedging against these risks in AI not only protects against vulnerabilities and accidents, but may also increase trust in and use of beneficial AI-enabled technologies. Increasing leaders' trust in AI may increase their willingness to explore beneficial applications of the technology and affect rates of diffusion. The level of trust in AI will impact how AI will alter the balance of power.[23] Relatedly, demonstrably safer AI may decrease domestic opposition to AI-enabled military technologies. Lin-Greenberg notes in his contribution that 74 percent of South Koreans and 72 percent of Germans are opposed to autonomous weapons, compared to 52 percent of Americans.[24] As he explains, the differences in public trust may result in unequal burden sharing, straining alliances.

Another key path for future success in the adoption of AI involves a focus on integration and ensuring compatibility and interoperability between systems, organizations, people, and even states. This is especially true as, according to experts including Cummings, AI is often most successful when used as a support tool, such as within command-and-control loops.

As Slayton emphasizes, the key element in AI success is actually a balanced human-machine dynamic: AI is most revolutionary in how it transforms human work through

[22] "Defending Against Adversarial Artificial Intelligence."

[23] Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power."

[24] Chris Deeney, "Six in Ten (61%) Respondents Across 26 Countries Oppose the Use of Lethal Autonomous Weapons Systems," IPSOS, last modified Jan. 21, 2019, https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons.

close interaction between humans and machines, "leveraging their distinctive kinds of intelligence," and "enabling new kinds of analysis and operations." Not by "replacing people." Humans must be educated and familiarized with AI and how it works, what its benefits are, and what its limitations are to fully realize AI's competitive advantage. Not only will focusing on AI training and education help further integrate AI systems, it will hedge against fears that advanced technologies will replace humans, as well as other risks such as overconfidence and brittleness.

Lin-Greenberg recommends something similar to ensure interoperability. Countries must socialize senior leaders and soldiers with AI-enabled capabilities via training and exercises, in particular through joint drills with alliance partners who might have different AI capabilities. Adopting standardized AI-integration methods such as labeling and formatting data, he continues, "will help streamline AI development and enhance interoperability, much in the same way that NATO standards today ensure that radios and other systems used by NATO partners can communicate with each other."

Finally, in order to fully enforce integration efforts and cohesion between states, humans, and machines, AI should be able to work effectively with existing technologies or legacy systems when possible. This will help facilitate adaptation and the adoption of technologies, as well as hedge against disparities among forces with different AI capabilities. In other words, successful AI adoption is about working together: New technologies working with old ones, allies working with one another to ensure interoperability, and militaries around the world working with AI in ways that bring out the strengths of both human and machine elements to increase international security.

Michael C. Horowitz *is professor of political science and interim director of Perry World House at the University of Pennsylvania.*

Lauren Kahn *is a research fellow at Perry World House.*

Christian Ruhl *is the program associate for Global Order at Perry World House.*

2. The AI that Wasn't There:

Global Order and the (Mis)Perception of Powerful AI

Mary (Missy) Cummings

Many preeminent thinkers and organizations have recently warned that advances in artificial intelligence (AI) could significantly shift the center of technology dominance from the United States to other, less democratic countries like China.[25] A fundamental issue with such discussions is the assumption that AI has advanced to the point of dramatically changing how militaries operate — which it may or may not have. However, it is not clear whether the achievement of such advances actually matters. It may be to a country's advantage to merely act as if it has advanced AI capabilities, something that is relatively easy to do. Such a pretense could then cause other countries to attempt to emulate potentially unachievable capabilities, at great effort and expense. Thus, the perception of AI prowess may be just as important as having such capabilities.

Before examining these issues, we must first ask, "What successes has AI had, both in commercial terms as well as for militaries worldwide?" To answer this question, we need a more precise definition of AI. The U.S. Department of Defense defines it as "the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical

[25] Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (New York: Houghton Mifflin Harcourt, 2018); and Sean Fleming, "World Order Is Going to Be Rocked by AI — This Is How," World Economic Forum, April 16, 2020, https://www.weforum.org/agenda/2020/02/ai-looks-set-to-disrupt-the-established-world-order-here-s-how/.

systems."[26]

The commercial world has seen qualified success in many aspects of AI, but the results have not been as transformative as projected. AI has dramatically improved voice recognition, which is now a mainstay of commercial businesses.[27] However, other successes have been more muted. For example, while computer vision has improved over the past 10 years, particularly for object recognition, the brittleness of underlying machine learning approaches has also become more evident over time.

Figure 1 illustrates the brittleness of computer vision using a deep-learning AI algorithm.

In this figure, there are three typical road vehicles — a school bus, a motor scooter, and a fire truck — each shown in a normal pose, as well as in three unusual poses (along with the probabilistic estimates of what the underlying computer vision algorithms see). Depending on which pose the vehicle was in, computer vision variously saw the items as a punching bag, a parachute, or a bobsled. These results demonstrate that this form of AI is unable to cope with different presentations of the same object. This is a well-known problem when it comes to driverless cars. Computer vision problems have been cited as contributing factors in many fatal Tesla crashes and the 2018 death of a pedestrian by an

[26] *Summary of the 2018 Department of Defense Artifical Intelligence Strategy*, Department of Defense, February 2019, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

[27] Erik Brynjolfsson and Andrew McAfee, "The Business of Artificial Intelligence," *Harvard Business Review*, 2017, https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence.

Uber self-driving car.[28]

AI has also been heralded as highly successful in playing games, specifically the TV game show Jeopardy! and the board games Go and chess.[29] While this may seem to be a breakthrough for AI-enabled decision-making, the reality is much more mundane. Such successes were achieved because the domains of games are deterministic, which means that the number of moves or the number of choices that can be made are *known*, albeit numerous. Computers excel over humans when searching a large space of known options. Where AI is decidedly much less capable is in drawing abstract conclusions that require judgment under conditions of uncertainty.[30] Indeed, Watson, the decision-making engine behind the Jeopardy AI success, has been deemed a general failure when it was extended to medical applications.[31] Alphabet's DeepMind medical AI is also facing increased scrutiny and skepticism.[32]

---

[28] Steve Crowe, "Tesla Autopilot Causes 2 More Accidents." *Robotics Trends*, July 12, 2016, https://www.roboticsbusinessreview.com/rbr/tesla_autopilot_causes_2_more_accidents/; Steve Lohr, "A Lesson of Tesla Crashes? Computer Vision Can't Do It All Yet," *New York Times*, Sept. 19, 2016, https://www.nytimes.com/2016/09/20/science/computer-vision-tesla-driverless-cars.html; and Troy Griggs and Daisuke Wakabayashi, "How a Self-Driving Uber Killed a Pedestrian in Arizona," *New York Times*, March 21, 2018, https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html.

[29] Trips Reddy, "Why It Matters that AI Is Better than Humans at Games Like Jeopardy," June 27, 2017, https://www.ibm.com/blogs/watson/2017/06/why-it-matters-that-ai-is-better-than-humans-at-their-own-games//.

[30] Mary "Missy" Cummings, "Man vs. Machine or Man + Machine?" *IEEE Intelligent Systems*, 29, no. 5 (September/October 2014): 62–69, https://doi.org/10.1109/MIS.2014.87.

[31] Eliza Strickland, "IBM Watson, Heal Thyself," *IEEE Spectrum* 56, no. 4 (April 2019): 24–31, https://doi.org/10.1109/MSPEC.2019.8678513.

[32] Donna Lu, "It's Too Soon to Tell if DeepMind's Medical AI Will Save Any Lives," *New Scientist*, July 31, 2019, https://www.newscientist.com/article/2212100-its-too-soon-to-tell-if-deepminds-medical-ai-will-save-any-lives/.

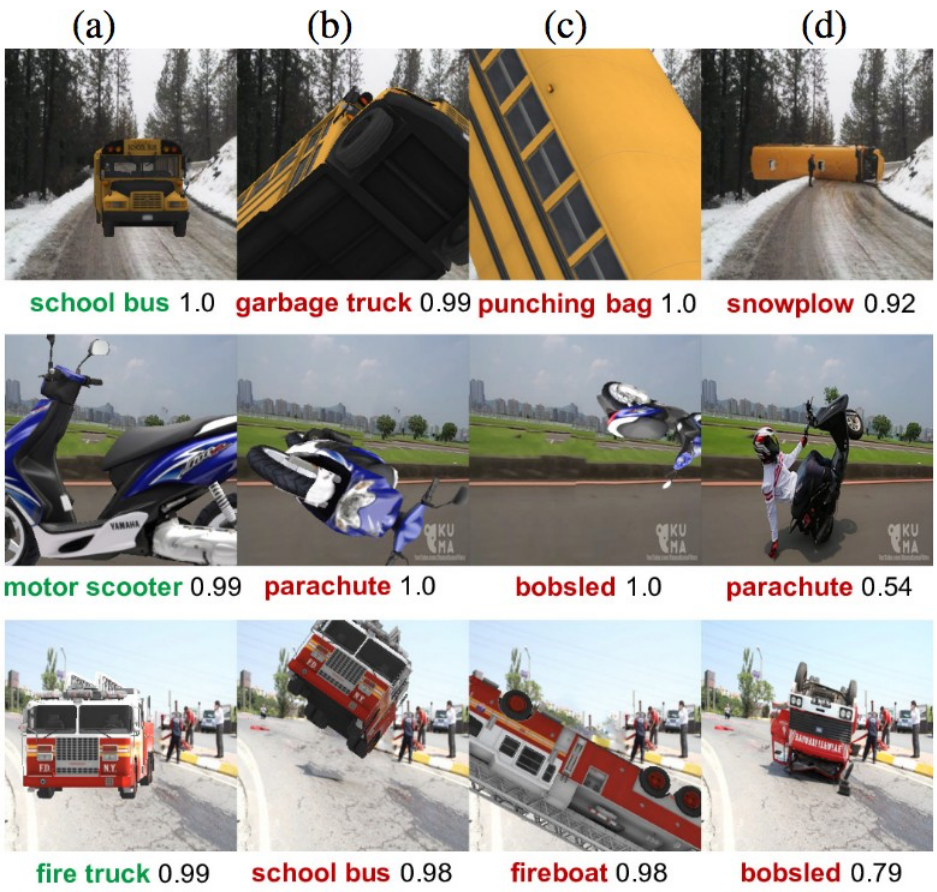

Figure 1: A deep-learning algorithm prediction for typical road vehicle poses in a 3D simulator (a) and for unusual poses (b-d). The computer's estimate of its probability of correctness follows the algorithm's label of what it thinks the object is.[33]

The inability of AI to handle uncertainty raises serious questions about how successful it will be in military settings. The fog of war is the paradigmatic example of uncertainty. Any AI-based system that has to reason about dynamic and uncertain environments is likely to be extremely unreliable, especially in situations never before encountered. Unfortunately, this is exactly the nature of warfare.

[33] Michael A. Alcorn et al., ("Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects," *arXiv*, November 2018, https://arxiv.org/abs/1811.11553.

How has AI fared in military settings thus far? It's difficult to tell, in part because there are very few actively deployed military systems that rely on AI. Drones — i.e., unmanned aerial vehicles — have advanced automated flight control systems, but they rely on rule-based coding to operate. This is a relatively simple and static algorithm and does not meet the definition of AI. The Tomahawk missile system, which is over 30 years old, uses primitive AI to match digital-camera scenes from its onboard camera to images in its database as it flies close to the earth.[34] While it is highly accurate, it cannot respond to dynamic scene changes and cannot cope with uncertainty.

The U.S. military is keen to use AI to improve automated target recognition in ways far more advanced than the Tomahawk missile. Such a capability would allow weapons systems to detect, identify, and destroy targets on their own in real time. While no military publishes exact statistics about such weapons systems, current reports suggest that little progress has been made in this area, likely due to issues with computer vision similar to those illustrated in Figure 1.[35]

Some consulting groups like Deloitte suggest that the best use of military AI lies not in weapons systems, but in support functions such as analyzing information. Analysis of intelligence — like satellite images or acoustic data — and logistics information may be an

---

[34] Erik V. Larson, *Technological Risk: The Case of the Tomahawk Cruise Missile* (Santa Monica, CA: Rand Corp., 1990).

[35] James A. Ratches, "Review of Current Aided/Automatic Target Acquisition Technology for Military Target Acquisition Tasks," *Optical Engineering* 50, no. 7 (2011), https://doi.org/10.1117/1.3601879; and Vincent Boulanin and Maaike Verbruggen, "Mapping the Development of Autonomy in Weapon Systems," Stockholm International Peace Research Institute, November 2017, https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems.

arena in which AI could improve planning processes.[36] Using AI as a support tool is quite a different approach from the AI-driven weapon-toting killer robots envisioned by some,[37] and suggests a more modest future for military AI applications.

The Illusion of AI

Despite the fact that AI has not been as successful in military and commercial settings as many people think, it is entirely possible that the perception of having all-powerful AI may be just as important as actually having it. A major factor driving the perception of who has the most advanced AI is who spends the most on it. Alphabet has spent more than $2 billion on DeepMind, which has a reputation as one of the most advanced AI companies in the world. However, DeepMind has produced very little in terms of revenue beyond successes in deterministic games like Alpha Go, calling into question DeepMind's supposed successes.[38]

---

[36] Frank Strickland, Joe Mariani, and Isaac Jenkins, "Military Readiness through AI: How Technology Advances Help Speed Up Our Defense Readiness," Deloitte Center for Government Insights, 2019, https://www2.deloitte.com/content/dam/insights/us/articles/5106_AI-for-govt-readiness/DI_AI-for-govt-readiness.pdf.

[37] "Arms: New Campaign to Stop Killer Robots." Human Rights Watch, April 23, 2013, https://www.cnas.org/the-ethical-autonomy-project-bibliography; and "Autonomous Weapons: an Open Letter from AI & Robotics Researchers," Future of Life Institute, 2015, http://futureoflife.org/open-letter-autonomous-weapons/.

[38] Gary Marcus, "DeepMind's Losses and the Future of Artificial Intelligence," *Wired*, Aug. 14, 2019, https://www.wired.com/story/deepminds-losses-future-artificial-intelligence/; and Julia Powles, "DeepMind's Latest A.I. Health Breakthrough Has Some Problems," *One Zero*, Aug. 6, 2019, https://onezero.medium.com/deepminds-latest-a-i-health-%20breakthrough-has-some-problems-5cd14e2c77ef.

The uncertain accomplishments of AI are important when it comes to the international arms race because there is serious concern that China is outpacing the United States in AI applications. But given the significant weaknesses of current AI development, it must be asked whether China is really ahead of the United States in AI development or if AI overhype and well-placed demonstrations have simply given the perception that China is ahead. If the latter, what are the ramifications of such a misperception?

The practice of claiming to possess all-powerful AI without actually having AI-driven systems is currently an issue in the commercial world of driverless cars. Companies developing driverless cars must rely on humans to significantly augment computer vision systems through data labelling: Humans must tell the car what it is seeing (road, bush, pedestrian, etc.), in the hope that after enough examples the car will "learn" these relationships on its own. As a result of the brittleness in such supervised approaches to learning, companies have not delivered on their promises of fleets of operational self-driving cars.[39] To date, no company has demonstrated the ability for sustained driving operations without a safety driver behind the wheel. This practice of "fake it till you make it" is well known in Silicon Valley and has shown up in other commercial settings, like when humans pretended to be calendar-scheduling chatbots or when call center employees acted as transcription AI for voice-to-text translation.[40]

[39] Roberto Baldwin, "Self-Driving Cars Are Taking Longer to Build than Everyone Thought," *Car and Driver*, May 10, 2020, https://www.caranddriver.com/features/a32266303/self-driving-cars-are-taking-longer-to-build-than-everyone-thought/.

[40] Olivia Solon, "The Rise of 'Pseudo-AI': How Tech Firms Quietly Use Humans to Do Bots' Work," *The Guardian*, July 15, 2018, https://mediawell.ssrc.org/2018/07/15/the-rise-of-pseudo-ai-how-tech-firms-quietly-use-humans-to-do-bots-work-the-guardian/.

The ramifications of this "fake it till you make it" culture in driverless cars has led to inflated and unrealistic expectations that are driving a hypercompetitive first-to-market race, which is quickly becoming prohibitively expensive. More than $100 billion has been spent on driverless car development,[41] with no end in sight due to problems in accuracy and reliability of computer vision. Because of spiraling costs, several company consolidations and partnerships have taken place in recent months and there is speculation that many companies will not survive.[42] The automotive industry's top investor at SoftBank has stated, "The risks are so big and opportunities so massive that there will be few players that have intellectual capital and financial capital."[43]

Investments in military AI are escalating similarly to those in commercial applications, fueling the concern that China may be outpacing the United States in military AI prowess. Indeed, just as America countered the Soviet Union's conventional military through outspending, particularly in terms of technological advancements, some fear China may be doing the same to the United States through investments in AI.[44]

---

[41] Paul A. Eisenstein, "Not Everyone Is Ready to Ride as Autonomous Vehicles Take to the Road in Ever-Increasing Numbers," *CNBC*, Oct. 14, 2018, https://www.cnbc.com/2018/10/14/self-driving-cars-take-to-the-road-but-not-everyone-is-ready-to-ride.html.

[42] Booke Masters, "Self-driving Car Companies Find that Going It Alone Is Difficult," *Financial Times*, July 1, 2019, https://www.ft.com/content/39c01b56-9be5-11e9-9c06-a4640c9feebb.

[43] Cory Weinberg, Aaron Tilley, and Kevin McLaughlin, "SoftBank's Ronen Says Self-Driving Market 'Big Boys' Game,'" *The Information*, June 13, 2019, https://www.theinformation.com/articles/softbanks-ronen-says-self-driving-market-big-boys-game.

[44] Robert O. Work and Greg Grant, "Beating the Americans at their Own Game: An Offset Strategy with Chinese Characteristics," Center for a New American Security, 2019, https://s3.amazonaws.com/files.cnas.org/documents/CNAS-Report-Work-Offset-final-B.pdf.

There is one critical difference in this historical comparison: Physical military systems can produce tangible illustrations of advancements, while claims of advanced AI are much harder to verify. For example, in the Cold War, the presence of ships operating on the high seas and visiting ports around the world communicated concrete, verifiable progress toward this goal. Claims of superior AI are much harder to verify since they are software-based. Moreover, as discussed above, it is not obvious in any given AI demonstration whether the results are real, or whether teams of humans are actually driving the success in a "Wizard of Oz" fashion.

Going forward, it is imperative for governments to monitor developments in military-related artificial intelligence, especially for weapons systems and in cyber security. However, it is equally important that they arm themselves with the capabilities to detect inflated or fake claims, so as not to invest large sums of money developing a counter-capability to a non- existent threat. Just as in ballistic-missile defense, where the Chinese use balloons to look like incoming threats to draw scarce counter-missile resources, humans must be able to detect whether AI is an actual threat in order to determine the timeliest and most cost-effective response.

Mary (Missy) Cummings *is a professor in the Duke University Electrical and Computer Engineering Department, and the Director of the Humans and Autonomy Laboratory. She is an American Institute of Aeronautics and Astronautics (AIAA) Fellow, and a member of the Defense Innovation Board. Her research interests include human supervisory control, explainable artificial intelligence, human-autonomous system collaboration, human-robot interaction, human-systems engineering, and the ethical and social impact of technology.*

3. Integrating Emerging Technology in Multinational Military Operations: The Case of Artificial Intelligence

Erik Lin-Greenberg

States acquire and develop new military technologies to gain an advantage on the battlefield, to increase the efficiency of operations, and to decrease operational risk. Although these technologies can help shift the balance of power, militaries often have difficulty integrating new equipment and practices into their existing force structure due to a combination of technical and institutional barriers.[45] The challenge of integrating new technologies into military operations is magnified in the case of alliances and other multinational coalitions. While members of these entities share security interests, each pursues its own national interests and has its own set of military priorities and procedures. As a result, each state may have different views on how and when to use these new technologies, complicating both planning and operations.

Multinational operations have long posed challenges to national security practitioners, but the development of more advanced military technologies has the potential to make cooperation even more complex.[46] NATO, for instance, faced significant challenges during the 1999 Kosovo Air War because the encrypted radio systems used by some member states could not communicate with those of other states.[47] More recently, allies have

[45] Michael C. Horowitz, *The Diffusion of Military Power: Causes and Consequences for International Politics* (Princeton, NJ: Princeton University Press, 2010).

[46] Eric Larson et al., *Interoperability: A Continuing Challenge in Coalition Air Operations* (Santa Monica, CA: Rand, 2000), 28–33.

[47] Larson et al., 20.

disagreed over the best policies and principles to guide cyber operations.[48] To explore the challenges of integrating new technologies into multinational military operations, I examine the case of artificial intelligence (AI) — a technology that is rapidly finding its way into military systems around the world. I argue that AI will complicate multinational military operations in two key ways. First, it will pose unique challenges to interoperability — the ability of forces from different states to operate alongside each other. Second, AI may strain existing alliance decision-making processes, which are frequently characterized by time consuming consultations between member states. These challenges, however, are not insurmountable. Just as allies have integrated advanced technologies such as GPS and nuclear weapons into operations and planning, they will also be able to integrate AI.

Conceptualizing Multinational Operations and Artificial Intelligence

Before tackling these issues, I outline some key definitions and concepts. Alliances and coalitions are cooperative endeavors in which members contribute resources in pursuit of shared security interests.[49] Alliances are typically more deeply formalized institutions whose operations are codified in treaties, while coalitions are typically shorter-term pacts created to achieve specific tasks — such as the defeat of an adversary. In today's security environment, states often carry out military operations alongside allies or coalition partners. These multinational efforts yield both political and operational benefits.

---

[48] Max Smeets, "Cyber Command's Strategy Risks Friction With Allies," *Lawfare*, May 28, 2019, https://www.lawfareblog.com/cyber-commands-strategy-risks-friction-allies; Christopher Porter and Klara Jordan, "Don't Let Cyber Attribution Debates Tear Apart the NATO Alliance," *Lawfare*, Feb. 14, 2019, https://www.lawfareblog.com/dont-let-cyber-attribution-debates-tear-apart-nato-alliance.

[49] Stephen Walt, *The Origins of Alliance* (Ithaca, NY: Cornell University Press, 1990); Glenn H. Snyder, *Alliance Politics* (Ithaca, NY: Cornell University Press, 1997).

Politically, multilateral operations can lend greater legitimacy to the use of force than unilateral actions. Since multinational efforts require the buy in of multiple states, they can help signal that a military operation is justified.[50] Alliances and coalitions also allow for burden sharing, with various member states each contributing to the planning and conduct of military operations. This can reduce the strain on any one state's military during contingency operations and allow states to leverage the specialized skills of different alliance or coalition members.

Despite the virtues of alliance and coalition operations, they also pose obstacles to strategic and operational coordination. Even if allies share security interests, they may have difficulty agreeing how to pursue their objectives — a task that becomes increasingly more challenging as the number of states in an alliance or coalition increases, and if the terms of the alliance commitment are vague (as they often are, to prevent allies from being drawn into conflicts they would prefer to avoid).[51] At a more operational level, allies and partners may face challenges when operating alongside each other because of technical, cultural, or procedural factors.[52] AI has the potential to exacerbate all of these issues.

---

[50] A large body of international relations scholarship suggests that support from multinational organizations can increase public support for the use of force and the legitimacy of military operations. See, Terrence L. Chapman, "Audience Beliefs and International Organization Legitimacy," *International Organization* 63, no. 4 (October 2009): 733–64, https://doi.org/10.1017/S0020818309990154; Terrence L. Chapman and Dan Reiter, "The United Nations Security Council and the Rally 'Round the Flag Effect," *Journal of Conflict Resolution* 48, no. 6 (December 2004): 886–909, https://www.jstor.org/stable/4149799.

[51] Michael Beckley, "The Myth of Entangling Alliances: Reassessing the Security Risks of U.S. Defense Pacts," *International Security* 39, no. 4 (Spring 2015): 7–48, https://doi.org/10.1162/ISEC_a_00197.

[52] *Joint Publication 3-16: Multinational Operations*, Joint Chiefs of Staff, March 1, 2019, I–3, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_16.pdf?ver=2019-11-14-170112-293.

In the military domain, AI has been increasingly used in roles that traditionally required human intelligence. In some cases, AI is employed as part of analytical processes, like the use of machine learning to help classify geospatial or signals intelligence targets. Or, it can be part of the software used to operate physical systems, like self-driving vehicles or aircraft. States around the world have already fielded a range of military systems that rely on AI technology. The U.S. Department of Defense, for instance, launched Project Maven to develop AI to process and exploit the massive volume of video collected by reconnaissance drones.[53] Similarly, Australia is working with Boeing to develop an advanced autonomous drone intended for use on combat missions, and the U.S. Navy is exploring the use of self-operating ships for anti-submarine warfare operations.[54] Military decision-makers look to these systems as ways of increasing the efficiency and reducing the risk of conducting military operations. Automating processes like signals analysis can reduce manpower requirements, while replacing sailors or soldiers with computers on the front lines can mitigate the political risks associated with suffering friendly casualties.[55]

---

[53] Theresa Hitchens, "In 1st Interview, PDUSDI Bingen Talks Artificial Intelligence, Project Maven, Ethics," *Breaking Defense*, Aug. 26, 2019, https://breakingdefense.com/2019/08/in-1st-interview-pdusdi-bingen-talks-artificial-intelligence-project-maven-ethics/.

[54] Ewen Levick, "Boeing's Autonomous Fighter Jet Will Fly Over the Australian Outback," *IEEE Spectrum*, 2 January 2020, https://spectrum.ieee.org/aerospace/military/boeings-autonomous-fighter-jet-will-fly-over-the-australian-outback; Megan Eckstein, "Sea Hunter USV Will Operate With Carrier Strike Group, As SURFDEVRON Plans Hefty Testing Schedule," *U.S. Naval Institute News*, Jan. 21, 2020, https://news.usni.org/2020/01/21/sea-hunter-usv-will-operate-with-carrier-strike-group-as-surfdevron-plans-hefty-testing-schedule.

[55] James Igoe Walsh and Marcus Schulzke, *Drones and Support for the Use of Force* (Ann Arbor: University of Michigan Press, 2018); Smriti Srivastava, "Indian Army Encourages Potent AI Mechanism to Reduce Manpower Dependency," *Analytics Insight*, Sept. 26, 2019, https://www.analyticsinsight.net/indian-army-encourages-potent-ai-mechanism-reduce-manpower-dependency/.

As states develop AI capabilities, leaders must consider the challenges that may arise when fielding AI as part of broader alliance or coalition efforts. First, alliance leaders must consider the unequal rates at which alliance members will adopt AI — and the consequences this could have on alliance and coalition operations. Second, leaders must consider how AI will affect two important components of alliance dynamics: shared decision-making and interoperability.

Challenges of Adopting AI[56]

New technology does not diffuse across the world at the same rate, meaning that some states will possess and effectively operate AI-enabled capabilities, while others will not.[57] This unequal distribution of technology can result from variation in material and human resources, or from political resistance to adoption. In the case of AI, large and wealthy states (e.g., the United States) and smaller, but technologically advanced countries (i.e., Singapore) have established robust AI development programs.[58] In contrast, less wealthy allies have tended to allocate their limited defense funding to other, more basic capabilities. Many of NATO's poorer members, for example, have opted to invest in

[56] For a deeper discussion of this argument, see, Erik Lin-Greenberg, "Allies and Artificial Intelligence: Obstacles to Operations and Decision-making," *Texas National Security Review* 3, no. 2 (Spring 2020).

[57] On the diffusion of new technologies, see, Everett M. Rogers, *Diffusion of Innovations*, 5th ed. (New York: Free Press, 2003); and Horowitz, *The Diffusion of Military Power*.

[58] Prashanth Parameswaran, "What's in the New US-Singapore Artificial Intelligence Defense Partnership?" *The Diplomat*, July 1, 2019, https://thediplomat.com/2019/07/whats-in-the-new-us-singapore-artificial-intelligence-defense-partnership/.

modernizing conventional equipment rather than developing new military AI capabilities.[59]

In addition to variation in material resources, public support for the development of military AI capabilities varies significantly across states, potentially shaping whether and how states develop AI. Even though AI enables a range of military capabilities, the notion of AI-enabled weapons often conjures up images of killer robots in the minds of the public. One recent survey finds strong opposition to the use of autonomous weapons among the population of U.S. allies like South Korea and Germany.[60] The public and many political and military decision-makers in these countries remain reluctant to delegate life-or-death decisions to computers, and worry about the implications of AI-enabled technologies making mistakes.[61] This type of resistance can lead states to ban the use of AI-enabled systems or hamper the development of AI technologies for military use, at least temporarily, as it did when Google terminated its involvement in Project Maven after employee protests.[62] The resulting divergence in capabilities between AI haves and have-

[59] "Modernization of the Armed Forces," Republic of Albania: Ministry of Defense, Oct. 12, 2019, http://www.mod.gov.al/eng/index.php/security-policies/others-from-mod/modernization/68-modernization-of-the-armed-forces.

[60] "Six in Ten (61%) Respondents Across 26 Countries Oppose the Use of Lethal Autonomous Weapons Systems," Ipsos, Jan. 21, 2019, https://www.ipsos.com/en-us/news-polls/human-rights-watch-six-in-ten-oppose-autonomous-weapons.

[61] Colin Clark, "Air Combat Commander Doesn't Trust Project Maven's Artificial Intelligence — Yet," *Breaking Defense*, Aug. 21, 2019, https://breakingdefense.com/2019/08/air-combat-commander-doesnt-trust-project-mavens-artificial-intelligence-yet/.

[62] "Country Views on Killer Robots," Campaign to Stop Killer Robots, Aug. 21, 2019, https://www.stopkillerrobots.org/wp-content/uploads/2019/08/KRC_CountryViews21Aug2019.pdf; Daisuke Wakabayashi and Scott Shane, "Google Will Not Renew Pentagon Contract that Upset Employees," *New*

nots within a multinational coalition may stymie burden-sharing. States without AI-enabled capabilities may be less able to contribute to missions, forcing better-equipped allies to take on a greater share of work — possibly generating friction between coalition members.

Challenges to Multinational Decision-making and Interoperability

Even if allies and coalition partners overcome the domestic obstacles to developing AI-enabled military technology, the use of these systems may still complicate decision-making and pose vexing interoperability challenges for multinational coalitions. These challenges can hamper multinational operations and potentially jeopardize cohesion among security partners.

Decision-making among allies is often characterized as a complex coordination game. Although allies share some set of objectives and goals, each state still maintains its own national interests. The negotiations needed to compromise on these divergent political interests can result in drawn-out decision-making timelines.[63] AI, however, has the potential to greatly accelerate warfare to what former U.S. Deputy Secretary of Defense Bob Work referred to as "machine speed."[64] The faster rate at which information is

York Times, June 1, 2018, https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html.

[63] Kenneth N Waltz, *Theory of International Politics* (Boston, MA: McGraw Hill, 1979), 163–70; John J. Mearsheimer, "The False Promise of International Institutions," *International Security* 19, no. 3 (Winter 1994/1995 1994): 32, https://www.jstor.org/stable/2539078.

[64] Robert Work, "Remarks to the Association of the U.S. Army Annual Convention," U.S. Department of Defense, Oct. 4, 2016, https://www.defense.gov/Newsroom/Speeches/Speech/Article/974075/remarks-to-the-association-of-the-us-army-annual-convention/.

produced and operations are carried out may strain existing alliance decision-making processes. The current NATO decision-making construct, for instance, requires the 29-member North Atlantic Council to debate and vote on issues related to the use of force.[65] As AI accelerates the speed of war, decision-making timelines may be compressed. Coalition leaders may find themselves making decisions without the luxury of extended debates.

At the more tactical level, the increased deployment of AI-enabled systems has the potential to complicate interoperability between coalition forces. Interoperability — "the ability to act together coherently, effectively, and efficiently to achieve tactical, operational, and strategic objectives" and "the condition achieved among communications-electronics systems…when information or services can be exchanged directly and satisfactorily between them and/or their users" — is critical to multinational operations.[66] Interoperability ensures military personnel and assets from each member state are equipped with both the technology and procedures that allow them to support other member states on the battlefield.

As new AI-enabled systems are introduced to the battlefield, they must — like older generations of technology — be able to communicate and integrate with each other and with existing legacy systems. The data-intensive underpinnings of AI, however, can make this a complicated task for political and technical reasons. Politically, states may be

[65] "North Atlantic Council," NATO, last updated Oct. 10, 2017, http://www.nato.int/cps/en/natohq/topics_49763.htm.

[66] *DOD Dictionary of Military and Associated Terms*, The Joint Staff, last updated January 2020, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf.

hesitant to share military and intelligence data even with close allies.[67] They may fear that providing unfettered access to data risks disclosing sensitive sources and methods, or revealing that states have been snooping on their allies.[68] These revelations could cause mistrust, strain political relationships, or compromise ongoing intelligence operations.

Even if allies are willing to share data, significant technical obstacles remain. Data produced by different states that could be used to train AI systems, for example, may be stored in different formats or tagged with different labels, making it difficult to integrate data from multiple states.[69] Further, much of this military and intelligence data resides on classified national networks that are not typically designed to enable easy information sharing. These information stovepipes have hampered past attempts at information and data sharing during coalition operations. This will only become more pronounced as data requirements increase in an age of AI-enabled warfare.[70]

---

[67] James Igoe Walsh, *The International Politics of Intelligence Sharing* (New York: Columbia University Press, 2009).

[68] Matthew Karnitschnig, "NSA Flap Strains Ties with Europe," *Wall Street Journal*, Feb. 9, 2014, https://www.wsj.com/articles/wave-of-nsa-reports-strain-ties-with-europe-1391971428?tesla=y.

[69] These data problems plague state militaries and would be magnified in the multinational context. See, Sydney J. Freedberg, "Pentagon's AI Problem Is 'Dirty' Data: Lt. Gen. Shanahan," *Breaking Defense*, Nov. 13, 2019, https://breakingdefense.com/2019/11/exclusive-pentagons-ai-problem-is-dirty-data-lt-gen-shanahan/.

[70] Matt Pottinger, Michael T. Flynn, and Paul D. Batchelor, *Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan* (Washington, DC: Center for New American Security, 2010), https://www.cnas.org/publications/reports/fixing-intel-a-blueprint-for-making-intelligence-relevant.

Preparing for "Machine Speed" Multinational Operations

Modern military operations are increasingly integrating AI, and potential rivals are developing robust AI capabilities. To deter these adversaries and to more efficiently carry out coalition operations, the United States must work with its allies to responsibly develop AI capabilities that are interoperable and support coalition decision-making processes. To these ends there are several steps the United States and its allies can take to better posture their forces for AI-enabled multinational operations.

First, allies and security partners should establish AI collaboration agreements that outline where and how AI will be used. Singapore and the United States, for instance, launched a partnership to coordinate on AI use in the national security domain.[71] These agreements would provide guidelines that would help develop shared tactics, techniques, and standardization procedures that would allow allies and coalition partners to more effectively integrate their AI-enabled capabilities. Coalition-wide standards on labeling and formatting of data, for instance, would help streamline AI development and enhance interoperability, much in the same way that NATO standards today ensure that radios and other systems used by NATO partners can communicate with each other.

Second, coalition leaders should explore how to streamline decision-making processes so they are responsive to AI-enabled warfare. This might include developing pre-established rules of engagement that delegate authorities to frontline commanders, or even to AI-enabled computers and weapon systems — something that leaders may be hesitant to do. Third, alliances should look to develop technologies and processes that overcome barriers to the sharing of sensitive data. To do this, allies could draw from previous agreements

[71] Parameswaran, "What's in the New US-Singapore Artificial Intelligence Defense Partnership?"

that governed the sharing of extremely sensitive intelligence.[72] Partners could also leverage procedures like secure multiparty computation, a privacy-preserving technique in which AI analyzes data inputs from sources that seek to keep the data secret, but produces outputs that are public to authorized users.[73]

Finally, allies and partners should help acclimate their forces to AI-enabled operations. For instance, multinational exercises might prominently feature AI-enabled capabilities like drone swarms or intelligence reports produced by AI-powered systems. Military leaders might also be asked to employ their own AI-enabled capabilities and to deter or defeat those of rivals during wargames. These exercises will help national security practitioners better understand what AI can and cannot do. This deeper understanding of AI will help decision-makers develop more informed plans and strategies, and help ensure multinational coalitions are ready for modern, AI-enabled warfare.

Erik Lin-Greenberg *is a postdoctoral fellow at the University of Pennsylvania's Perry World House. In Fall 2020, he will start as an assistant professor of political science at MIT.*

[72] Walsh, *The International Politics of Intelligence Sharing*.

[73] Andrew C. Yao, "Protocols for Secure Computations," in *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, SFCS '82 (Washington, DC, USA: IEEE Computer Society, 1982), 160–64.

4. The Militarization of Artificial Intelligence

Paul Scharre

Militaries are racing to adopt artificial intelligence (AI) with the aim of gaining military advantage over competitors. And yet, there is little understanding of AI's long-term impact on warfare.[74] The current wave of interest in AI, largely driven by the deep-learning revolution, is relatively new, and machine-learning technology continues to evolve rapidly, buoyed by exponential growth in data and computing power. Society is only beginning to grapple with the myriad opportunities for and challenges in applying deep learning to practical problems outside of the research lab.[75] Understanding AI's effect on warfare is particularly challenging given the sometimes-secretive nature of military technology development and the lack of everyday settings in which military AI systems can be incrementally tested.

---

[74] Tim Dutton, "An Overview of National AI Strategies," *Medium*, June 28, 2018, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd. 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, Nov. 28, 2017, https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power.

[75] For a non-technical 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, July 2018, https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know; 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.

Academics, policymakers, and military practitioners have begun tackling how AI may transform warfare, contributing to a burgeoning literature in recent years.[76] This article presents three lenses through which one may examine AI and warfare: specific applications of AI; historical analogies; and the technical characteristics of AI systems today. Each approach has merits and is worthy of further study.

[76] Keir A. Lieber and Daryl G. Press, "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence," *International Security* 41, no. 4 (Spring 2017): 9–49, https://doi.org/10.1162/ISEC_a_00273; Austin Long and Brendan Rittenhouse Green, "Stalking the Secure Second Strike: Intelligence, Counterforce, and Nuclear Strategy," *Journal of Strategic Studies* 38, no. 1–2 (2015): 38–73, https://doi.org/10.1080/01402390.2014.958150; Edward Geist and Andrew J. Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War," RAND Corporation, 2018, https://www.rand.org/pubs/perspectives/PE296.html; "Artificial Intelligence and the Military: Forever Altering Strategic Stability," Technology for Global Security, Feb. 13, 2019, https://www.tech4gs.org/uploads/1/1/1/5/111521085/ai_and_the_military_forever_altering_strategic_stability__t4gs_research_paper.pdf; Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 3, no 1. (May 2018): 38–57, https://doi.org/10.15781/T2639KP49; Kenneth Payne, "Artificial Intelligence: A Revolution in Strategic Affairs?" *Survival* 60, no. 5, (2018): 7–32, https://doi.org/10.1080/00396338.2018.1518374; Kenneth Payne, "Artificial Intelligence Is About to Revolutionise Warfare. Be Afraid," *New Scientist,* Sept. 12, 2018, https://www.newscientist.com/article/mg23931950-600-artificial-intelligence-is-about-to-revolutionise-warfare-be-afraid/; Kenneth Payne, *Strategy, Evolution, and War: From Apes to Artificial Intelligence* (Washington, DC: Georgetown University Press, 2018); Miles Brundage et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," *arXiv*, February 2018, https://arxiv.org/abs/1802.07228; and Gregory C. Allen and Taniel Chan, "Artificial Intelligence and National Security," Belfer Center, July 2017, https://www.belfercenter.org/sites/default/files/files/publication/AI%20NatSec%20-%20final.pdf.

Specific Applications of AI

One method for attempting to understand how AI may affect warfare is to examine specific applications of AI that could be consequential. For example, there are a variety of AI applications — from early warning to delivery systems or counter-force options — that could affect nuclear stability and are deserving of exploration. Scholars have begun considering the implications of high-consequence uses of AI, including nuclear operations, lethal autonomous weapons, crisis stability, and swarm combat.[77] There are many other AI applications that ought to be studied to better comprehend how AI will change missile defense, undersea warfare, cyber operations, and counter-terrorism. One challenge to this approach is that military capabilities are in a constant process of innovating on both offense and defense, capabilities and countermeasures. Analysts must guard against the tendency to envision only one application of AI without also

[77] Michael C. Horowitz, "When Speed Kills: Lethal Autonomous Weapons Systems, Deterrence and Stability," *Journal of Strategic Studies* 42, no. 6): 764–88, https://doi.org/10.1080/01402390.2019.1621174; Vincent Boulanin and Maaike Verbruggen, "Mapping the Development of Autonomy in Weapon Systems," Stockholm International Peace Research Institute, November 2017, https://www.sipri.org/publications/2017/other-publications/mapping-development-autonomy-weapon-systems; Jean-Marc Rickli, "Some Considerations of the Impact of LAWS on International Security: Strategic Stability, Non-State Actors and Future Prospects," Paper Presented to the Meeting of Experts on Lethal Autonomous Weapons Systems at the Convention on Certain Conventional Weapons (CCW), April 16, 2015, Geneva Switzerland, 2, http://www.unog.ch/80256EDD006B8954/(httpAssets)/B6E6B974512402BEC1257E2E0036AAF1/$file/2015_LAWS_MX_Rickli_Corr.pdf; Jean-Marc Rickli, "The Impact of Autonomous Weapons Systems on International Security and Strategic Stability," in *Defence Future Technologies: What We See on the Horizon*, ed. Quentin Ladetto (Thun, Switzerland: Armasuisse, 2017), 62; and Jürgen Altmann and Frank Sauer, "Autonomous Weapon Systems and Strategic Stability," *Survival* 59, no. 5 (2017): 117–42, http://dx.doi.org/10.1080/00396338.2017.1375263.

considering the countermeasures that AI (or other technologies) may enable. For example, autonomous systems could be used for swarming attacks, but swarms could also be used defensively. Given that AI is a general-purpose technology with a wide variety of potential uses, there are many issues worth exploring, making this a fruitful area of research.

AI in History

A second lens through which to view the military adoption of AI is that of history, specifically, lessons from historical analogies. As a general-purpose technology, AI is more like computers, electricity, or the internal combustion engine than a discrete technology such as locomotives or fighter planes. AI's impact on warfare is likely to be transformative and vast, unfolding over many decades. One valuable, historical analogy to the AI revolution is the first and second industrial revolutions. Just as the industrial revolutions unleashed a broad process of industrialization across society, today's AI revolution is spurring a process of "cognitization."[78] The industrial revolutions led to the creation of machines that were stronger than humans for certain tasks, offloading physical labor to machines. Similarly, AI enables the creation of special-purpose machines that are more "intelligent" than humans for specific tasks, offloading *cognitive* labor to machines. As AI is applied to a variety of tasks, it is likely to have sweeping effects on society. One estimate assesses that nearly half the tasks currently done in the United States could be automated using existing technology, including routine physical and cognitive labor

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

across a wide swath of occupations.[79] Similar widespread applications are likely in warfare. The net effect of the cognitization of warfare could be changes as significant as those brought about by the industrial revolutions and mechanized warfare, which included the adoption of locomotives, machine guns, airplanes, submarines, tanks, and trucks.

The challenge for scholars is to estimate the net effect of these myriad changes in the cognitization of warfare. The industrial revolutions enabled the expansion of warfare to new domains (undersea and air) and enabled a vast increase in destructive capacity. AI affects the cognition — or in military terms, command-and-control — of military systems. Machine intelligence can then be used to enable autonomous systems or tools to aid human decision-making. Robots are one application of AI, and some scientists have predicted a "Cambrian explosion" of robots of various shapes and sizes,[80] which could have many military uses. However, the most significant effects of AI are likely to be on command and control. In computer games such as Starcraft or Dota 2, AI systems have been able to achieve superhuman performance using the same basic elements as human players but with better command and control.[81] AI systems could enable military forces to operate faster, more cohesively, and with greater precision and coordination than humans alone can. The result could be to accelerate the pace of battle beyond human decision-

[79] Michael Chui, James Manyika, and Mehdi Miremadi, "The Four Fundamentals of Workplace Automation," *McKinsey Quarterly*, November 2015, https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/four-fundamentals-of-workplace-automation.

[80] Gill Pratt, "Is a Cambrian Explosion Coming for Robotics?" *Journal of Economic Perspectives* 29, no. 3 (Summer 2015): 51–60, https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.29.3.51.

[81] "OpenAI Five," OpenAI, https://openai.com/five/; "AlphaStar: Mastering the Real-Time Strategy Game Starcraft II," DeepMind, Jan. 24, 2019, https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii.

making — what some Chinese scholars have called a "battlefield singularity" or what some Western scholars have labelled "hyperwar."[82] Such a development would be profoundly dangerous, unleashing forces on the battlefield that would be beyond human control, at least for some period of time. Yet, competitive pressures could drive such an arms race in speed.

The Technical Characteristics of AI Today

The third lens through which to view the effects of AI on warfare is with regard to the specific nature of AI technology today and what makes it different from other emerging or existing dual-use technologies. It is striking that thousands of AI researchers have spoken out about the dangers of the militarization of AI or an "AI arms race."[83] Militaries routinely adopt new technologies to improve their capabilities, yet the militarization of AI in particular has proven controversial in a way that the military adoption of computers or networking, for example, has not. While there are no doubt many factors behind these concerns, it is worth asking: Do AI scientists know something about AI technology that policymakers don't? The answer is quite clearly "yes." Current AI and machine-learning

[82] 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; John R. Allen and Amir Husain, "On Hyperwar," *Proceedings* 143, no. 7 (July 2017), https://www.usni.org/magazines/proceedings/2017/july/hyperwar; and Amir Husain et al., *Hyperwar* (Austin, TX: SparkCognition Press, 2018).

[83] "Autonomous Weapons: An Open Letter From AI & Robotics Researchers," Future of Life Institute, http://futureoflife.org/open-letter-autonomous-weapons/.

methods are powerful but brittle.[84] AI systems can achieve superhuman performance in some settings, yet fail catastrophically in others. AI systems are also vulnerable to bias, adversarial attacks, data poisoning, reward hacking, and other types of failure.[85] In addition, machine-learning systems often exhibit surprising emergent behaviors, in ways both good and bad.[86] While military AI technical experts understand these flaws, they have yet to percolate to the minds of senior leaders, who often have only heard of the potential benefits of AI technology. Some caution is warranted. Deep-learning technologies are powerful, but are also insecure and unreliable. One AI scientist recently compared machine learning to "alchemy."[87] Perhaps policymakers would be more cautious if AI were presented to them as a kind of "militarizing alchemy."

---

[84] For an overview of the limitations of current narrow AI systems, see, Amodei et al., "Concrete Problems in AI Safety," Preprint, submitted July 25, 2016, 4, https://arxiv.org/pdf/1606.06565.pdf.

[85] Larry Hardesty, "Study Finds Gender and Skin-type Bias in Commercial Artificial-Intelligence Systems," *MIT News*, Feb. 11, 2018, http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212; Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women," *Reuters*, Oct. 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; Anh Nguyen, Jason Yosinski, and Jeff Clune, "Deep Neural Networks Are Easily Fooled: High Confidence Predictions for Unrecognizable Images," Computer Vision and Pattern Recognition Conference, 2015, https://arxiv.org/abs/1412.1897; James Vincent, "Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less than a Day," *The Verge*, May 24, 2016, https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist; and Nicolas Papernot et al., "Practical Black-Box Attacks against Machine Learning," Preprint, submitted March 19, 2017, https://arxiv.org/pdf/1602.02697.pdf.

[86] Dario Amodei and Jack Clark, "Faulty Reward Functions in the Wild," *OpenAI*, Dec. 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," Preprint, submitted March 8, 2018, 6, https://arxiv.org/pdf/1803.03453.pdf.

[87] Ali Rahimi, "NIPS 2017 Test-of-Time Award presentation," YouTube, https://www.youtube.com/watch?v=ORHFOnaEzPc; and Matthew Hutson, "AI Researchers Allege that

The risk is not that AI systems don't work, in which case they would be unlikely to be deployed at all, but that they work perfectly in training environments but fail spectacularly in wartime. Their brittle nature means that subtle changes in environment or the data they use could dramatically change their performance.[88] Wars are rare, which is a good thing, but the downside is that militaries may not have access to realistic datasets on which to train AI systems. Many military tasks can be accurately rehearsed in peacetime, such as aircraft takeoff and landing, aerial refueling, or driving vehicles. In these circumstances, AI systems can likely be designed, tested, and verified over time to achieve adequate levels of performance, in some cases better than humans. But adapting to enemy tactics is another matter entirely. For battlefield decision-making, not only at the operational level but also at the tactical level, humans will be required. The human

---

Machine Learning Is Alchemy," *Science*, May 3, 2018, https://www.sciencemag.org/news/2018/05/ai-researchers-allege-machine-learning-alchemy.

[88] See examples of accidents with Tesla's autopilot. Jim Puzzanghera, "Driver in Tesla Crash Relied Excessively on Autopilot, but Tesla Shares Some Blame, Federal Panel Finds," *Los Angeles Times*, Sept. 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, Sept. 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" 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.

mind remains the most advanced cognitive processing system on the planet. AI systems, for all their prowess, perform poorly at adapting to novel or unexpected situations, which abound in warfare. If militaries deploy AI systems before they are fully tested and verified in an attempt to stay ahead of competitors, they risk sparking a "race to the bottom" when it comes to AI safety.[89]

As militaries continue to pursue artificial intelligence, leaders should be aware of the significant risks that could come with its adoption. Many military applications of AI will be inconsequential, but others could be concerning, such as the use of AI in nuclear operations or lethal decision-making. The widespread adoption of AI could accelerate the pace of military operations, pushing warfare beyond human control, while the pursuit of AI capabilities risks a "race to the bottom" in AI safety. Militaries are exploring the benefits of AI, which are likely to be significant, but they should also study the potential risks that may emerge from military applications of AI, as well as how to mitigate those risks.

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.*

[89] 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.

5. The Promise and Risks of Artificial Intelligence: A Brief History

Rebecca Slayton

Artificial intelligence (AI) has recently become a focus of efforts to maintain and enhance U.S. military, political, and economic competitiveness. The Defense Department's 2018 strategy for AI, released not long after the creation of a new Joint Artificial Intelligence Center, proposes to accelerate the adoption of AI by fostering "a culture of experimentation and calculated risk taking," an approach drawn from the broader *National Defense Strategy*.[90] But what kinds of calculated risks might AI entail? The AI strategy has almost nothing to say about the risks *incurred* by the increased development and use of AI. On the contrary, the strategy proposes using AI to *reduce* risks, including those to "both deployed forces and civilians."[91]

While acknowledging the possibility that AI might be used in ways that reduce some risks, this brief essay outlines some of the risks that come with the increased development and deployment of AI, and what might be done to reduce those risks. At the outset, it must be acknowledged that the risks associated with AI cannot be reliably calculated. Instead, they are emergent properties arising from the "arbitrary complexity" of information systems.[92] Nonetheless, history provides some guidance on the kinds of

[90] *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, U.S. Department of Defense, 2018, 7, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

[91] *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 11.

[92] For the original concept of "arbitrary complexity," see, Frederick P. Brooks, Jr., "No Silver Bullet: Essence and Accidents of Software Engineering," *Computer* 20, no. 4 (April 1987): 10-19, https://doi.org/10.1109/MC.1987.1663532. For an application of this concept to a military context, see,

risks that are likely to arise, and how they might be mitigated. I argue that, perhaps counter-intuitively, using AI to manage and reduce risks will require the development of uniquely human and social capabilities.

A Brief History of AI, From Automation to Symbiosis

The Department of Defense strategy for AI contains at least two related but distinct conceptions of AI. The first focuses on *mimesis* — that is, designing machines that can mimic human work.[93] The strategy document defines mimesis as "the ability of machines to perform tasks that normally require human intelligence — for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action."[94] A somewhat distinct approach to AI focuses on what some have called human-machine *symbiosis*, wherein humans and machines work closely together, leveraging their distinctive kinds of intelligence to transform work processes and organization. This vision can also be found in the AI strategy, which aims to "use AI-enabled information, tools, and systems to empower, not replace, those who serve."[95]

Of course, mimesis and symbiosis are not mutually exclusive. Mimesis may be understood as a means to symbiosis, as suggested by the Defense Department's proposal to "augment the capabilities of our personnel by offloading tedious cognitive or physical

---

Rebecca Slayton, *Arguments that Count: Physics, Computing, and Missile Defense, 1949-2012* (Cambridge, MA: MIT Press, 2013).

[93] A. M. Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (October 1950): 433–60, http://www.jstor.org/stable/2251299.

[94] *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 5.

[95] *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 4.

tasks."[96] But symbiosis is arguably the more revolutionary of the two concepts and also, I argue, the key to understanding the risks associated with AI.

Both approaches to AI are quite old. Machines have been taking over tasks that otherwise require human intelligence for decades, if not centuries. In 1950, mathematician Alan Turing proposed that a machine can be said to "think" if it can persuasively imitate human behavior, and later in the decade computer engineers designed machines that could "learn." By 1959, one researcher concluded that "a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program."[97]

Meanwhile, others were beginning to advance a more interactive approach to machine intelligence. This vision was perhaps most prominently articulated by J.C.R. Licklider, a psychologist studying human-computer interactions. In a 1960 paper on "Man-Computer Symbiosis," Licklider chose to "avoid argument with (other) enthusiasts for artificial intelligence by conceding dominance in the distant future of cerebration to machines alone." However, he continued: "There will nevertheless be a fairly long interim during which the main intellectual advances will be made by men and computers working together in intimate association."[98]

Notions of symbiosis were influenced by experience with computers for the Semi-Automatic Ground Environment (SAGE), which gathered information from early warning

[96] *Summary of the 2018 Department of Defense Artificial Intelligence Strategy*, 7.

[97] A.L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal* 3, no. 3 (July 1959): 535, https://doi.org/10.1147/rd.33.0210.

[98] J.C.R. Licklider, "Man-Computer Symbiosis," *IRE Transactions on Human Factors in Electronics* HFE-1, no. 1 (March 1960): 4–11, https://doi.org/10.1109/THFE2.1960.4503259.

radars and coordinated a nationwide air defense system. Just as the Defense Department aims to use AI to keep pace with rapidly changing threats, SAGE was designed to counter the prospect of increasingly swift attacks on the United States, specifically low-flying bombers that could evade radar detection until they were very close to their targets. Unlike other computers of the 1950s, the SAGE computers could respond instantly to inputs by human operators. For example, operators could use a light gun to select an aircraft on the screen, thereby gathering information about the airplane's identification, speed, and direction. SAGE became the model for command-and-control systems throughout the U.S. military, including the Ballistic Missile Early Warning System, which was designed to counter an even faster-moving threat: intercontinental ballistic missiles, which could deliver their payload around the globe in just half an hour. We can still see the SAGE model today in systems such as the Patriot missile defense system, which is designed to destroy short-range missiles — those arriving with just a few minutes of notice.[99]

SAGE also inspired a new and more interactive approach to computing, not just within the Defense Department, but throughout the computing industry. Licklider advanced this vision after he became director of the Defense Department's Information Processing Technologies Office, within the Advanced Research Projects Agency, in 1962. Under Licklider's direction, the office funded a wide range of research projects that transformed how people would interact with computers, such as graphical user interfaces and computer networking that eventually led to the Internet.[100]

---

[99] For more on the history of SAGE and its influence on computing, see, Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1996); and Slayton, *Arguments that Count*.

[100] For more on Licklider and his influence, see, M. Mitchell Waldrop, *The Dream Machine: J.C.R. Licklider and the Revolution that Made Computing Personal* (New York: Viking Penguin, 2001).

The technologies of symbiosis have contributed to competitiveness not primarily by replacing people, but by enabling new kinds of analysis and operations. Interactive information and communications technologies have reshaped military operations, enabling more rapid coordination and changes in plans. They have also enabled new modes of commerce. And they created new opportunities for soft power as technologies such as personal computers, smart phones, and the Internet became more widely available around the world, where they were often seen as evidence of American progress. Mimesis and symbiosis come with somewhat distinct opportunities and risks. The focus on machines mimicking human behavior has prompted anxieties about, for example, whether the results produced by machine reasoning should be trusted more than results derived from human reasoning. Such concerns have spurred work on "explainable AI" — wherein machine outputs are accompanied by humanly comprehensible explanations for those outputs.

By contrast, symbiosis calls attention to the promises and risks of more intimate and complex entanglements of humans and machines. Achieving an optimal symbiosis requires more than well-designed technology. It also requires continual reflection upon and revision of the models that govern human-machine interactions. Humans use models to design AI algorithms and to select and construct the data used to train such systems. Human designers also inscribe models of use — assumptions about the competencies and preferences of users, and the physical and organizational contexts of use — into the technologies they create. Thus, "like a film script, technical objects define a framework of action together with the actors and the space in which they are supposed to act."[101]

[101] Madeleine Akrich, "The De-Scription of Technical Objects," in *Building Society: Studies in Sociotechnical Change*, ed. Wiebe E. Bijker and John Law (Cambridge, MA: MIT Press, 1992).

Scripts do not completely determine action, but they configure relationships between humans, organizations, and machines in ways that constrain and shape user behavior. Unfortunately, these interactively complex sociotechnical systems often exhibit emergent behavior that is contrary to the intentions of designers and users.

Competitive Advantages and Risks

Because models cannot adequately predict all of the possible outcomes of complex sociotechnical systems, increased reliance on intelligent machines leads to at least four kinds of risks: The models for how machines gather and process information, and the models of human-machine interaction, can both be inadvertently flawed or deliberately manipulated in ways not intended by designers. Examples of each of these kinds of risks can be found in past experiences with "smart" machines.

First, changing circumstances can render the models used to develop machine intelligence irrelevant. Thus, those models and the associated algorithms need constant maintenance and updating. For example, what is now the Patriot missile defense system was initially designed for air defense but was rapidly redesigned and deployed to Saudi Arabia and Israel to defend against short-range missiles during the 1991 Gulf War. As an air defense system it ran for just a few hours at a time, but as a missile defense system it ran for days without rebooting. In these new operating conditions, a timing error in the software became evident. On Feb. 25, 1991, this error caused the system to miss a missile that struck a U.S. Army barracks in Dhahran, Saudi Arabia, killing 28 American soldiers. A software patch to fix the error arrived in Dhahran a day too late.[102]

[102] "Patriot Missile Defense: Software Problem Led to System Failure at Dhahran, Saudi Arabia," General Accounting Office, Report to the Chairman, Subcommittee on Investigations and Oversight, Committee on

Second, the models upon which machines are designed to operate can be exploited for deceptive purposes. Consider, for example, Operation Igloo White, an effort to gather intelligence on and stop the movement of North Vietnamese supplies and troops in the late 1960s and early 1970s. The operation dropped sensors throughout the jungle, such as microphones, to detect voices and truck vibrations, as well as devices that could detect the ammonia odors from urine. These sensors sent signals to overflying aircraft, which in turn sent them to a SAGE-like surveillance center that could dispatch bombers. However, the program was a very expensive failure. One reason is that the sensors were susceptible to spoofing. For example, the North Vietnamese could send empty trucks to an area to send false intelligence about troop movements, or use animals to trigger urine sensors.[103]

Third, intelligent machines may be used to create scripts that enact narrowly instrumental forms of rationality, thereby undermining broader strategic objectives. For example, unpiloted aerial vehicle operators are tasked with using grainy video footage, electronic signals, and assumptions about what constitutes suspicious behavior to identify and then kill threatening actors, while minimizing collateral damage.[104] Operators following this script have, at times, assumed that a group of men with guns was planning an attack, when in fact they were on their way to a wedding in a region where celebratory gun firing is customary, and that families praying at dawn were jihadists rather than simply observant Muslims.[105] While it may be tempting to dub these mistakes "operator

---

science, Space, and Technology, House of Representatives, February 1992, https://www.gao.gov/assets/220/215614.pdf.

[103] Edwards, *The Closed World*.

[104] Hugh Gusterson, *Drone: Remote Control Warfare* (Cambridge, MA: MIT Press, 2016).

[105] See, for example, Lucy Draper, "The Wedding that Became a Funeral: U.S. Still Silent One Year on from Deadly Yemen Drone Strike," *Newsweek*, Dec. 12, 2014, https://www.newsweek.com/wedding-became-funeral-us-still-silent-one-year-deadly-yemen-drone-strike-291403. For an analysis of the process by which

errors," this would be too simple. Such operators are enrolled in a deeply flawed script — one that presumes that technology can be used to correctly identify threats across vast geographic, cultural, and interpersonal distances, and that the increased risk of killing innocent civilians is worth the increased protection offered to U.S. combatants. Operators cannot be expected to make perfectly reliable judgments across such distances, and it is unlikely that simply deploying the more precise technology that AI enthusiasts promise can bridge the very distances that remote systems were made to maintain. In an era where soft power is inextricable from military power, such potentially dehumanizing uses of information technology are not only ethically problematic, they are also likely to generate ill will and blowback.

Finally, the scripts that configure relationships between humans and intelligent machines may ultimately encourage humans to behave in machine-like ways that can be manipulated by others. This is perhaps most evident in the growing use of social bots and new social media to influence the behavior of citizens and voters. Bots can easily mimic humans on social media, in part because those technologies have already scripted the behavior of users, who must interact through liking, following, tagging, and so on. While influence operations exploit the cognitive biases shared by all humans, such as a tendency to interpret evidence in ways that confirm pre-existing beliefs, users who have developed machine-like habits — reactively liking, following, and otherwise interacting without reflection — are all the more easily manipulated. Remaining competitive in an age of AI-mediated disinformation requires the development of more deliberative and reflective modes of human-machine interaction.

---

operators make judgments, see, Hugh Gusterson, "Drone Warfare in Waziristan and the New Military Humanism," *Current Anthropology* 60, no. 19 (February 2019): S77–S86, https://doi.org/10.1086/701022.

Conclusion

Achieving military, economic, and political competitiveness in an age of AI will entail designing machines in ways that encourage humans to maintain and cultivate uniquely human kinds of intelligence, such as empathy, self-reflection, and outside-the-box thinking. It will also require continual maintenance of intelligent systems to ensure that the models used to create machine intelligence are not out of date. Models structure perception, thinking, and learning, whether by humans or machines. But the ability to question and re-evaluate these assumptions is the prerogative and the responsibility of the human, not the machine.

Rebecca Slayton *is an associate professor in the Science & Technology Studies Department and the Judith Reppy Institute of Peace and Conflict Studies, both at Cornell University. She is currently working on a book about the history of cyber security expertise.*