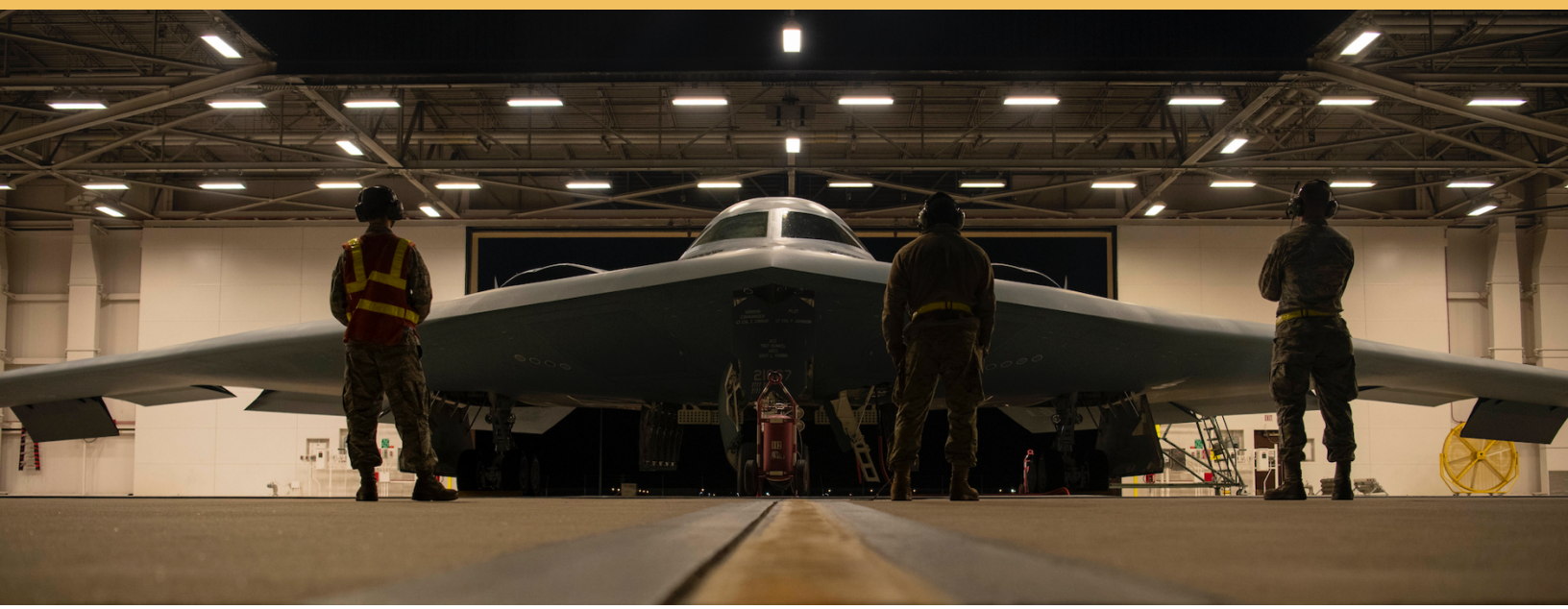
International Security at the NUCLEAR NEXUS

PROJECT ON NUCLEAR ISSUES



DECEMBER 17, 2020

Left of Launch

ARTIFICIAL INTELLIGENCE AT THE NUCLEAR NEXUS

By **Lindsey Sheppard**

Popular media and policy-oriented discussions on the incorporation of artificial intelligence (AI) into nuclear weapons systems frequently focus on matters of launch authority—that is, whether AI, especially machine learning (ML) capabilities, should be incorporated into the decision to use nuclear weapons and thereby reduce the role of human control in the decisionmaking process. This is a future we should avoid. Yet while the extreme case of automating nuclear weapons use is high stakes, and thus existential to get right, there are many other areas of potential AI adoption into the nuclear enterprise that require assessment. Moreover, as the conventional military moves rapidly to adopt AI tools in a host of mission areas, the overlapping consequences for the nuclear mission space, including in nuclear command, control, and communications (NC3), may be underappreciated.[1] AI may be

Banner Image: *Crew chiefs of the 509th Aircraft Maintenance Squadron B-2 Spirit look on during preflight checks in support of an annual command and control exercise designed to train U.S. Strategic Command forces and assess joint operation readiness, at Whiteman Air Force Base, Missouri, Oct. 22, 2020.* Credit: Staff Sgt. Dylan Nuckolls, U.S. Air Force

used in ways that do not directly involve or are not immediately recognizable to senior decisionmakers. These areas of AI application are far left of an operational decision or decision to launch and include four priority sectors: (1) security and defense; (2) intelligence activities and indications and warning; (3) modeling and simulation, optimization, and data analytics; and (4) logistics and maintenance. Given the rapid pace of development, even if algorithms are not used to launch nuclear weapons, ML could shape the design of the next-generation ballistic missile or be embedded in the underlying logistics infrastructure. ML vision models may undergird the intelligence process that detects the movement of adversary mobile missile launchers and optimize the tipping and queuing of overhead surveillance assets, even as a human decisionmaker remains firmly in the loop in any ultimate decisions about nuclear use. Understanding and navigating these developments in the context of nuclear deterrence and the understanding of escalation risks will require the analytical attention of the nuclear community and likely the adoption of risk management approaches, especially where the exclusion of AI is not reasonable or feasible.

For any of these applications, deploying AI and ML requires time to experiment and iterate with solutions and approaches. The nuclear weapons community's culture of zero tolerance for mistakes will be in tension with the culture of experimentation and bottom-up innovation that has allowed for AI successes in other national security applications. AI and ML applications are not without their technical risk and quirks that must be uncovered during the technology development and acquisition processes. This paper lays out a framework for understanding areas where the utilization of AI could impact the nuclear and strategic deterrence missions. As Philip Reiner and Alexa Wehsener write, "[t]he question is not *whether* AI be integrated into NC3, but rather *where*, *to what extent*, and *at what risk*."[2] Finally, this paper seeks to address how the integration of AI into these areas might impact strategic stability. While the insertion of ML into nuclear launch authority is widely considered highly problematic, the risks associated with the insertion of AI and ML into other components of the NC3 system, especially those with dual-use conventional/nuclear applications, may present significantly underappreciated risks. Even in these areas of application, the relationships between AI capability and strategic effects are sometimes counterintuitive and are maturing with AI technology itself, requiring further scholarship that brings together the AI and nuclear communities.

AI Primer

AI, in its most basic forms and applications, has existed for decades. Despite the explosion of headlines and academic articles touting the impressive results of AI in every field from advertising to virtual aerial combat, AI actually emerged as an academic field of study in the United States in the 1950s.[3] Since then, it has grown as an umbrella term to encompass a range of subdisciplines within the field: ML, natural language processing, computer vision, knowledge representation, automated reasoning, and robotics. Through the decades, the underlying approaches to creating computers that act or think humanly or apply human rationality to actions and decisions

---

[1] *The Future of Defense Task Force Report*, released by the United States House Armed Services Committee in September 2020, calls for a Manhattan Project-style focus on AI for defense. The report calls for Congress and the DOD to evaluate AI options for all Major Defense Acquisition Programs (MDAPs), to require all new MDAPs to be "AI-ready," and to expand DOD authorities to bring in new technologies. United States House Armed Services Committee, *Future of Defense Task Force Report* (Washington, DC: September 29, 2020), 7, https://armedservices.house.gov/2020/9/future-of-defense-task-force-releases-final-report.

[2] Philip Reiner and Alexa Wehsener, "The Real Value of Artificial Intelligence in Nuclear Command and Control," War on the Rocks, November 4, 2019, https://warontherocks.com/2019/11/the-real-value-of-artificial-intelligence-in-nuclear-command-and-control/.

[3] Patrick Tucker, "An AI Just Beat a Human F-16 Pilot In a Dogfight — Again," Defense One, August 20, 2020, https://www.defenseone.com/technology/2020/08/ai-just-beat-human-f-16-pilot-dogfight-again/167872/.

(depending on the philosophy of what constitutes "true" artificial intelligence) have also evolved through the life of the field of study.

This current period of AI excitement is primarily driven by the application of ML to a variety of problems and disciplines following research developments demonstrating unexpectedly positive performance at image recognition and classification in 2012.[4] ML is the process of using models, or algorithms, to make useful predictions or decisions using data relevant to a defined problem or task. Using ML, systems can improve performance over time as model performance adapts to new inputs without a human explicitly programming a new rule or "if -> then" heuristic.[5]

The application of machine learning in uncontrolled or real-world environments remains experimental and iterative in nature.

While ML has existed as a subdiscipline in AI almost as long as AI has been a field of study, it largely resided in academic and research communities. However, today it seems that AI is everywhere, with applications ranging from medicine to advertising to secretive defense programs. As seen through commercial applications, many people are familiar with AI through daily use. ML allows for systems to tailor and improve performance to user preferences over time—think of email spam filters or the robotic assistants on mobile devices that responds to voice commands. This broad use is due to the fact that ML excels at solving certain types of problems that exist in almost every field: anomaly detection, classification, clustering, optimization, data generation, ranking, and recommendation.[6] However, the application of ML in uncontrolled or real-world environments remains experimental and iterative in nature.[7] Continued high-profile examples of biased ML applications demonstrate the experimental nature of real-world ML.[8] Anxieties around the use of AI are founded in the knowledge that contextually-rich and uncontrolled environments will stress and strain ML systems, perhaps in ways that would have severe consequences. The correct approach, then, for those looking to incorporate ML into systems and processes, is to identify uses that lend themselves to controlled settings and timelines conducive to experimentation and iteration.

The successes of ML in recent years are due in part to the confluence of trends in growing amounts of data, the availability of relatively inexpensive computing power, and the miniaturization of microelectronic components for

---

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems 1, December 2012, 1097–1105, http://www.cs.toronto.edu/~hinton/absps/imagenet.pdf.

[5] For a general audience seeking to familiarize with machine learning concepts and terms, two good starting places are: "Machine learning," Wikipedia, https://en.wikipedia.org/wiki/Machine_learning; and Vincent Boulanin, "Artificial Intelligence: A primer," in *The Impact Of Artificial Intelligence On Strategic Stability And Nuclear Risk Volume I Euro-Atlantic Perspectives*, Vincent Boulanin, ed. (Stockholm, Sweden: SIPRI, 2019), https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf.

[6] For definitions of these problem types, see Michael Chui et al., *Notes from the AI Frontier: Insights from Hundreds of Use Cases* (Washington, DC: McKinsey Global Institute, April 2018), 5, https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20ai%20frontier%20applications%20and%20value%20of%20deep%20learning/notes-from-the-ai-frontier-insights-from-hundreds-of-use-cases-discussion-paper.ashx.

[7] Email correspondence with Ashley Llorens, August 16, 2020.

[8] Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, October 10, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G; and Chaim Gatenberg, "Twitter plans to change how image cropping works following concerns over racial bias," The Verge, October 2, 2020, https://www.theverge.com/2020/10/2/21498619/twitter-image-cropping-update-racial-bias-machine-learning.

computers and devices. Applying ML to new or existing processes requires the algorithmic model that calculates the statistical relationships between inputs and outputs, usable data that captures the things about the world the algorithm must "learn" or generalize, the computing power and network infrastructure necessary to process the model and data, and finally the right human expertise and talent. The phases of this process include "data collection, data pre-processing, building datasets, model training and refinement, evaluation, and deployment to production."[9] Depending on the approach taken, each component may vary in degree, but in general all are needed. AI is often computationally intensive, requiring significantly more computing power than a traditional individual computer workstation can provide and requiring organizations to make software and computing infrastructure choices throughout an enterprise that are compatible with data and computing requirements.[10] Some applications of AI even require special microprocessors, or chips, specifically designed for ML computing. Further, the process of preparing usable data sets is time and resource intensive.

Regardless of application, AI systems are far from fool proof. ML introduces its own vulnerabilities, performance considerations, and quirks. As ML is incorporated into system software, new avenues for access and exploit are available, many of which may be unintuitive when compared to non-ML systems. A 2019 Microsoft study introduces two categories of ML failures: **intentional failures** that result from the active efforts of adversarial actors and **unintentional failures** that are technically correct from the viewpoint of ML functionality but that are incorrect from the viewpoint of the application.[11] Intentional AI failures may also be referred to as **adversarial AI**, a term that describes the practice of exploiting both AI and ML through some hacking mechanism (e.g., the example of placing a sticker on a stop sign to fool an image classifier) to achieve a desired outcome. A related term, **adversarial machine learning**, is used within the ML community to specifically mean "malicious inputs designed to fool machine learning models."[12] As Kurakin, Goodfellow, and Bengio state, "[i]t has been shown that machine learning models are often vulnerable to adversarial manipulation of their input intended to cause incorrect classification (Dalvi et al., 2004). In particular, neural networks and many other categories of machine learning models are highly vulnerable to attacks based on small modifications of the input to the model at test time (Biggio et al., 2013; Szegedy et al., 2014; Goodfellow et al., 2014; Papernot et al., 2016b)."[13] ML may also be used to generate those malicious inputs used in adversarial machine learning attacks, though it is not necessary. It should be noted that **adversarial machine learning** is sometimes used imprecisely to refer broadly to an adversarial use of ML, that is, the use of ML against another system that itself may or may not include ML models.

Given this exploitable vulnerability, those looking to apply ML must consider the security of their systems. Unfortunately for those in safety critical fields, ML security is a nascent field of study where it seems attackers may have the advantage for the foreseeable future. Further, the current reliance on large data sets requires that data protection and security also be considered, as adversaries may target the data required to train an ML model. For example, data poisoning is one such way that attackers can exploit the necessity of training data for ML solutions.

---

[9] For a brief overview of the necessary steps to apply machine learning, see: "Machine Learning Workflow: Streamlining Your ML Pipeline," run:ai, https://www.run.ai/guides/machine-learning-operations/machine-learning-workflow/#:~:text=Machine%20learning%20workflows%20define%20which,evaluation%2C%20and%20deployment%20to%20production.

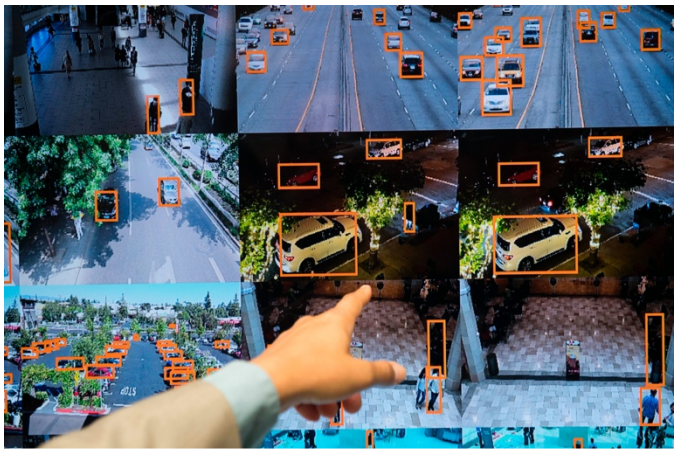
[10] Ben Buchanan, "The U.S. Has AI Competition All Wrong: Computing Power, Not Data, Is the Secret to Tech Dominance," *Foreign Affairs*, August 7, 2020, https://www.foreignaffairs.com/articles/united-states/2020-08-07/us-has-ai-competition-all-wrong.

[11] Ram Shankar Siva Kumar, Jeffrey Snover, David O'Brien, Kendra Albert, and Salome Viljoen, "Failure Modes in Machine Learning," Microsoft, November 2019, https://docs.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning.

[12] Alexey Kurakin, Ian Goodfellow, and Samy Bengio, "Adversarial Machine Learning at Scale," arXiv e-print, February 11, 2017, https://arxiv.org/pdf/1611.01236.pdf.

[13] Ibid.

By injecting malicious samples into the training data set, attackers can impact or manipulate model performance. This method of attack is not necessarily sophisticated; as Steinhardt et al. state in their paper on defending against data poisoning attacks, "for a system trained on user data, an attacker can inject malicious data simply by creating a user account."[14] Additional vulnerabilities exist when systems rely on open-source libraries for training data, as attackers similarly have access.[15] Particularly for safety critical applications, various research communities continue to study mechanisms and processes to ensure, verify, and validate that large volumes of data remain secure and untampered.



A display shows a vehicle and person recognition system intended for use by law enforcement.
Source: SAUL LOEB/AFP via Getty Images

The leading edge of AI may be seen most clearly in demonstrations from both academic and private sector research hubs. OpenAI, for example, has demonstrated impressive results on natural language processing through its General Pre-trained Transformer (GPT) progression of language models that generate AI-written text that is difficult to distinguish from human-written text.[16] However the leading edge of research development and demonstration often looks flashier than the leading edge of AI application. Comparatively, the state of scalable and feasible AI applications is easily accessible ML computer vision or language application programming interfaces (API) that may be deployed for tasks such as image recognition or automatic machine translation of text or audio.[17]

Caution is warranted, however. The current surge in research and development is reminiscent of previous cycles in the AI field. AI literature often references the two previous winters, or periods of significant decline in research, funding, and general expectations, following periods of significant hype and promise. The first winter, through the 1970s, began with a decline in the late-1960s. It followed the years of promise in "good, old-fashioned artificial intelligence," a rules-based approach where humans define all of the rules governing a program's behavior, as exemplified by the creation of computers that play checkers. The second winter began in the late-1980s to early-1990s following the promise of expert systems, an approach driven by attempts to implement in computers the problem-solving knowledge of domain-specific human experts.[18] The AI systems that came out of these periods—robotic process automation, symbolic programming, and expert systems—would likely not be considered AI today, though they are fairly mature in most cases. Interestingly enough, the promise of expert systems in the 1980s led to many discussions that would feel all too familiar today. Much like conversations on present ML capability, expert

---

[14] Jacob Steinhardt, Pang Wei Koh, and Percy Living, "Certified Defenses for Data Poisoning Attacks," 31st Conference on Neural Information Processing Systems, 2017, https://papers.nips.cc/paper/2017/file/9d7311ba459f9e45ed746755a32dcd11-Paper.pdf.

[15] An example is seen in the cybersecurity community, where models may rely on open-source training data sets, allowing adversaries to train their attacking systems on the same data. Hyrum S. Anderson and Phil Roth, "EMBER: An Open Dataset for Training Static PE Malware Machine Learning Models," arXiv e-print, April 16, 2018, https://arxiv.org/pdf/1804.04637.pdf.

[16] At the time of publication, GPT-3 family of models was the most recent capability demonstration and publication from the OpenAI team. See: https://arxiv.org/abs/2005.14165.

[17] Sean Maday, "Aircraft Visrecce with Computer Vision," LinkedIn, January 3, 2019, https://www.linkedin.com/pulse/aircraft-visrecce-computer-vision-sean-maday/.

[18] Colin Garvey, "Broken Promises & Empty Threats: The Evolution Of Ai In The USA, 1956-1996," *Technology's Stories* 6, no. 1 (March 2018), http://www.technologystories.org/ai-evolution/.

systems can "address only very narrow areas of expertise and have limited capability to encode common sense."[19] These systems are brittle; they "cannot determine if the problem they are trying to solve is within their 'area of expertise.'"[20] Systems are domain specific and not readily transferable. Questions remain on handling uncertainty and how best to present information to human users. There "are genuine problems that must be addressed and solved before AI can be used successfully in the complex world of the present."[21] "The solution is to use our 'clever' expert systems to solve the mundane and allow the human to free up [their] time . . . to solve the problems that the machine cannot solve."[22] Time will tell if the cycle will repeat for AI and ML, resulting in a third winter. It is arguable that the broad applicability of ML to a variety of business sectors and potential adopters may stave off such an outcome. Those looking to apply AI and ML, however, must balance the potential with the limitations and capability of the technology today.

AI Risks and Opportunities at the Nuclear Nexus

Nuclear weapons technology as a field is noted for its conservative and slow approach to integrating new technologies.[23] This caution is understandable, given the significant potential consequences of technical failures, and requires sober assessments of the reliability, predictability, and suitability of any new technology in support of the nuclear and strategic deterrence mission. Outside of this space, the commercial sector continues to move rapidly toward the development and application of AI and ML, with many parts of the conventional military enterprise in hot pursuit. It is essential that decisionmakers carefully evaluate the capabilities of AI and ML in the context of the unique risks that exist within the nuclear and strategic deterrence mission while appreciating the broad intersections between the nuclear enterprise and the rest of the U.S. military in areas where AI adoption may be seen to hold considerable promise and in which comparative risks may not be fully understood. Four areas stand out as requiring a more integrated approach: (1) security and defense; (2) intelligence activities and indications and warning; (3) modeling and simulation, optimization, and data analytics; and (4) logistics and maintenance. These areas are "everyday" functions of the nuclear mission, where data may be plentiful and tasks may be defined with clear objectives in relatively controlled environments. The problem types, timelines, and conditions necessary for AI and ML adoption and experimentation exist in these four functional areas. For example, given a set of operational objectives, requirements, and constraints, defense planners and analysts use ML and optimization techniques to analyze weapons and resource allocation.

Nevertheless, each of these left of launch areas may conceal escalatory risks. It is not unreasonable to expect that the use of AI may result in improved system capability and resilience. Such an improvement may simultaneously result in benefits to the United States and fundamental shifts in the dynamics between nuclear armed states. These risks are underappreciated and warrant further exploration. While technologists may be more inclined to

---

[19] Jeffrey Melaragno and Mary Kay Allen, *A Plan for the Application of Artificial Intelligence to DoD Logistics* (Bethesda, MA: Logistics Management Institute, October 1989), https://apps.dtic.mil/dtic/tr/fulltext/u2/a216066.pdf.

[20] Jude E. Franklin, Cora Lackey Carmody, Karl Keller, Todd S. Levitt, and Brandon L. Buteau, "Expert System Technology for the Military: Selected Samples," *Proceedings of The IEEE* 76, no. 10 (October 1988), https://www.researchgate.net/publication/236377474.

[21] Ibid.

[22] Ibid.

[23] Vincent Boulanin, "AI & Global Governance: AI and Nuclear Weapons – Promise and Perils of AI for Nuclear Stability," United Nations University Centre for Policy Research, December 7, 2018, https://cpr.unu.edu/ai-global-governance-ai-and-nuclear-weapons-promise-and-perils-of-ai-for-nuclear-stability.html.

incorporate AI and ML, there is a need for greater attention and focus by a range of experts in both the nuclear and technology development communities.

Security and Defense

The nuclear enterprise must secure and defend its installations, nuclear weapons, and systems from both cyber and physical intrusions. Commercial sector entities are turning to AI and ML for continuous monitoring, anomaly detection, and evaluation of physical or cybersecurity risks as part of their overall security posture. For physical installation security applications, existing applications of AI build upon existing electro-optical or video surveillance systems for monitoring and detection by adding an automatic identification and queuing component. Perimeter security through "intelligent video" approaches deploy computer vision, ML, and data analytics to physical security solutions to surveil areas and detect incidents in order to queue a response.[24] The private sector is also turning to ML for cyber defense. Currently, ML is deployed within the product offerings of cybersecurity firms such as Cylance, FireEye, and Crowdstrike.[25] Cybersecurity applications use ML for signature-based activities, such as event detection, malware detection and classification, automation, orchestration, user activity monitoring, network mapping, detection of lateral movement inside the network, automated incident response, and attack mitigation.

However, decisions by any actors, including the nuclear community, to incorporate AI and ML must be made with the understanding that new technology may introduce new vulnerabilities into systems. For example, a purely video or photo system leaves the system vulnerable to adversarial inputs designed to fool computer vision systems, as seen in demonstrations of glasses, patches, or t-shirts that cause ML surveillance systems to misclassify human wearers.[26] In cybersecurity products, ML algorithms are "living models" that may be exploited by attackers to gain access to a target system, requiring patching or model retraining.[27] Decisionmakers overseeing near-term modernization efforts must be prepared to evaluate the benefits and risks of incorporating ML into security and defense systems, particularly if they turn to third-party

Decisions by any actors, including the nuclear community, to incorporate artificial intelligence and machine learning must be made with the understanding that new technology may introduce new vulnerabilities into systems.

[24] John Merlino, "Getting smarter about perimeter and border protection with advanced analytics," Axis Communications, February 15, 2018, https://www.axis.com/blog/secure-insights/perimeter-border-protection-advanced-analytics/.

[25] John P. Mello Jr., "When machine learning is hacked: 4 lessons from Cylance," TechBeacon, August 22, 2019, https://techbeacon.com/security/when-machine-learning-hacked-4-lessons-cylance; David Krisiloff, "Churning Out Machine Learning Models: Handling Changes in Model Predictions," FireEye, April 9, 2019, https://www.fireeye.com/blog/threat-research/2019/04/churning-out-machine-learning-models-handling-changes-in-model-predictions.html; and "Crowdstrike Introduces Enhanced Endpoint Machine Learning Capabilities and Advanced Endpoint Protection Modules," CrowdStrike, February 13, 2017, https://www.crowdstrike.com/resources/news/crowdstrike-introduces-enhanced-endpoint-machine-learning-capabilities-and-advanced-endpoint-protection-modules/.

[26] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter, "A General Framework for Adversarial Examples with Objectives," arXiv e-print, April 4, 2019, https://arxiv.org/pdf/1801.00349.pdf.

[27] "Resolution for BlackBerry Cylance Bypass," BlackBerry ThreatVector Blog, July 21, 2019, https://blogs.blackberry.com/en/2019/07/resolution-for-blackberry-cylance-bypass.

vendors or seek to mirror commercial capability to secure assets in the nuclear enterprise.

The nuclear community must also keep an eye on the technical horizon in securing physical installations and cyber assets for both capability and threats. Considering the pace of technology development and current efforts to clear airspace for small form factor drones such as quadcopters, one could imagine a system that automatically responds to an intrusion through a passive response, such as persistent on-target surveillance until a human team arrived, providing a capability to augment physical security teams that are tasked with securing large installations that house missiles and other sensitive facilities. With the volumes of data collected through surveillance systems, pattern of life analysis around the installations is a possibility. Such a capability could help security teams identify anomalous behavior or threat behavior, perhaps from foreign intelligence, that would not meet the threshold for tripping a perimeter breach detection system.

The threat space will also continue to advance. In the cyber domain, nuclear network defenders must contend with non-ML cyber threats that are, at present, far more effective and common than ML-enabled attacks while preparing for a future of ML threats.[28] While largely academic or focused on proof-of-concept demonstration, researchers are exploring the efficacy of ML to automate attacks against and probe and evade defenses of both traditional static and ML-enabled cybersecurity systems. For example, publicly available research documents use adversarial machine learning against malware classifier engines that classify software as malicious or benign to gain access to target systems.[29] Even if the risks of incorporating ML are assessed to be too great, the nuclear community may soon live in a world where its assets and systems are targeted with the assistance of ML.

Intelligence Activities and Indications and Warning

The advent of readily available commercial satellite imagery, miniaturization of satellite and imagery technology, ML, and computer vision are revolutionizing the geospatial intelligence (GEOINT) and overhead imagery analysis fields. Publicly accessible overhead imagery is significantly higher in resolution from the imagery available only to governments a few decades ago.[30] Satellites and the imagery equipment on them have drastically reduced in size, weight, and power consumption, while the fidelity and quality of images produced improves. While national assets have similarly improved and by all indications remain superior in quality, the commercial satellite industry has made available to average users a large quantity of high-quality and timely overhead imagery that surpasses that of national assets a short time ago.[31] Corporations, commercial retailers, and stock investors are now turning to real-time and near-real-time satellite imagery to monitor facilities, predict corporate profits (e.g., counting cars in parking lots), and inform investment decisions.[32] AI tools play a growing role in targeting collection assets, assessing and collating enormous data sets, and validating data and detecting anomalies.

---

[28] Ben Buchanan, John Bansemer, Dakota Cary, Jack Lucas, Micah Mussar, *Automating Cyber Attacks: Hype and Reality* (Washington, DC: Center for Security and Emerging Technology, November 2020), https://cset.georgetown.edu/wp-content/uploads/CSET-Automating-Cyber-Attacks.pdf.

[29] Octavian Suciu, Scott E. Coull, and Jeffrey Johns, "Exploring Adversarial Examples in Malware Detection," arXiv e-print, April 13, 2019, https://arxiv.org/pdf/1810.08280.pdf; and Kim Zetter, "Researchers Easily Trick Cylance's AI-Based Antivirus Into Thinking Malware Is 'Goodware'," Vice, July 18, 2019, https://www.vice.com/en_us/article/9kxp83/researchers-easily-trick-cylances-ai-based-antivirus-into-thinking-malware-is-goodware.

[30] Cade Metz, "'Businesses Will Not Be Able to Hide': Spy Satellites May Give Edge From Above," *New York Times*, January 24, 2019, https://www.nytimes.com/2019/01/24/technology/satellites-artificial-intelligence.html.

[31] Geoff Brumfiel, "Trump Tweets Sensitive Surveillance Image of Iran," NPR, August 30, 2019, https://www.npr.org/2019/08/30/755994591/president-trump-tweets-sensitive-surveillance-image-of-iran.

[32] Frank Partnoy, "Stock Picks from Space," *The Atlantic,* May 2019, https://www.theatlantic.com/magazine/archive/2019/05/stock-value-satellite-images-investing/586009/.

The nuclear warning and analysis system is likely to rely on many of these capabilities. For the nuclear community, this progression of commercial application of ML mirrors the foundational intelligence work that occurs over years to piece together a picture of activity in places of interest and requires ML and computer vision to process of the troves of imagery data now available. In fact, the agencies and organizations of the U.S. Intelligence Community are currently working toward digital modernization and exploring the use of AI and ML to support their missions.[33] One example of the potential use of AI and ML is the targeting of adversary nuclear assets. Nations such as Russia, China, and the Democratic Peoples' Republic of Korea (DPRK) are investing in ground-mobile nuclear assets, complicating counterforce targeting strategies. Possessing rail or road-mobile nuclear forces can increase their survivability by hindering an adversary's ability to locate, track, and target them. Given the breadth of search area, mobile target search and threat detection data collection efforts can benefit from an analytic and data-driven approach to optimize sensor placement. Intelligence analysts may use AI in the data processing necessary to detect patterns of movement, alert status, and maintenance schedules, to name a few. AI may also open the possibility for new analysis, such as determining pattern of life movements with Chinese mobile missile launchers.

This nexus of AI and NC3 exemplifies the need for continued exploration of the relationships between technology application and strategic effects. Research from Rebecca Hersman et al. explores the counterintuitive impacts of increased situational awareness enabled by modern technologies on crisis stability and escalation pathways afforded by modern technologies, such as AI-enabled intelligence, reconnaissance, and surveillance.[34] Keir Lieber and Darryl Press write about the disruption of strategic stability through counterforce targeting.[35] With an increased ability to find, fix, and track assets of interest, dynamic tasking of surveillance assets enables a greater speed and precision of targeting of either nuclear or conventional assets. Such a capability could disrupt strategic stability by increasing first-strike incentives or the perception of first-strike opportunity and undermining secure second-strike assurance. However, the persistence and completeness of geographic coverage and timelines for intelligence processing, exploitation, and dissemination required to hold the necessary number of assets at risk to impact strategic stability may not be realistically achievable. While modernization efforts should use AI, ML, and optimization techniques to more efficiently task overhead assets, place sensors, and sift through the volumes of data now available to the strategic intelligence mission, additional scholarship is warranted to better understand this potential stability inflection point.

Modeling and Simulation, Optimization, and Data Analytics

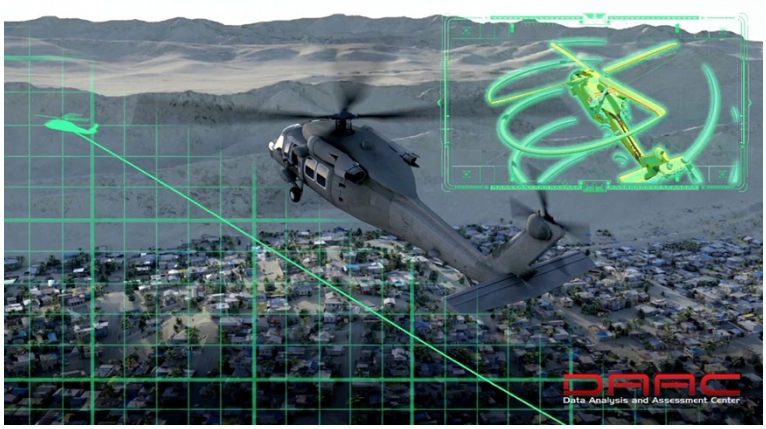
AI and ML bring additional tools to the toolbox to tackle the myriad optimization and analytics problems inherent to the nuclear and strategic deterrence missions. While these areas may not be directly captured in NC3, AI is most present in this area of application given the computational resource and technical expertise resident at national labs and research facilities. Both real-world and computational approaches to modeling and simulation allow humans to better understand specific phenomena. Wargaming tabletop exercises use modeling and simulation to explore behavioral dynamics and decisionmaking in the lead up to and during crises. Computational modeling and

---

[33] For example, see "CIA Achieves Key Milestone in Agency-Wide Modernization Initiative," U.S. Central Intelligence Agency, Press Release, October 1, 2015, https://www.cia.gov/news-information/press-releases-statements/2015-press-releases-statements/cia-achieves-key-milestone-in-agency-wide-modernization-initiative.html.

[34] Rebecca Hersman, Reja Younis, Bryce Farabaugh, Bethany Goldblum, and Andrew Reddie, *Under the Nuclear Shadow: Situational Awareness Technology and Crisis Decisionmaking* (Washington, DC: CSIS, March 2020), https://ontheradar.csis.org/analysis/final-report/.

[35] 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://www.belfercenter.org/publication/new-era-counterforce-technological-change-and-future-nuclear-deterrence.



The ERDC Data Analysis and Assessment Center provides a visualization of Helios simulations of maneuvering rotocraft. The Engineering Resilient Systems program uses Helios high-fidelity simulations to inform acquisition decisions.
Source: U.S. Army

simulation, including ML and optimization techniques, are used during system design, concept exploration, and testing. As a supplement to real-world testing, physics-based models can simulate the behavior of existing systems in critical but often difficult-to-recreate operating environments, such as certain failure modes in aircraft.

In both computational and real-world simulation, AI often models strategies or designs humans may not have otherwise tried, as seen in the example of AlphaGo developing new strategies and moves in the game of Go.[36] Indeed, researchers at DeepMind applied AI to scientific discovery and announced in late 2020 that they had used AI to solve a 50-year-old challenge in the biology field.[37] The use of generative adversarial networks (GANs) and reinforcement learning for wargaming could lead to different perspectives on where to deploy forces and where to target adversary forces, providing the nudge for humans to think differently. Though the use of AI for system design optimization is a technique that has been developed over decades in academia and research laboratories, it has gained renewed attention by organizations such as DARPA and the military services.[38] While traditional design techniques rely on best practice and human intuition to generate system design alternatives, some conventional military research for the U.S. Department of Defense (DOD) and U.S. Army uses tradespace exploration—a computational modeling technique that may use AI—and optimization to evaluate resiliency in system design.[39]

For the nuclear community, two areas that may see the application of AI-aided design are survivable and resilient space architectures and next-generation weapons systems design. Commercial space companies use computational techniques to optimize the myriad factors, such as the number and orbits of satellites and the number and location of ground stations, against a set of operational objectives and constraints, such as desired coverage and cost. Within

---

[36] David Silver and Demis Hassabis, "AlphaGo Zero: Starting from scratch," DeepMind, October 18, 2017, https://deepmind.com/blog/article/alphago-zero-starting-scratch.

[37] AlphaFold Team, "AlphaFold: a solution to a 50-year-old grand challenge in biology," DeepMind, November 30, 2020, https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology.

[38] "Evolving Computers from Tools to Partners in Cyber-Physical System Design," DARPA, August 2, 2019, https://www.darpa.mil/news-events/2019-08-02. "Decision-support" is one such areas where the nuclear and engineering communities use the same phrase with distinct meanings. The nuclear community uses "decision support" to mean the conferencing process during crisis. More closely aligned with the this paper, the modeling and simulation community uses "decision-support" to mean analytic processes or computational tools that process and present data and analysis to aid decisionmakers, often in relation to a specific system design or process definition problem.

[39] "Tradespace exploration" is a computational modeling approach that often incorporates artificial neural networks (a form of ML) to quantify and define relationships among system design alternatives, system attributes, and design variables so that decisionmakers may evaluate trade-offs across designs, operational characteristics, and capabilities. For examples of use in engineering decision-making, see: Valerie B. Sitterle et al., "Systems Engineering Resiliency: Guiding Tradespace Exploration within an Engineered Resilient Systems Context," *Procedia Computer Science* 44, (2015): 649–658, doi:10.1016/j.procs.2015.03.013; Valerie B. Sitterle et al., "Integrated Toolset and Workflow for Tradespace Analytics in Systems Engineering," 24th Annual International Council on Systems Engineering (INCOSE) Symposium, Las Vegas, NV, June 30-July 3, 2014, doi:10.1002/j.2334-5837.2014.tb03153.x.

the defense industrial base, developers of next-generation missiles use trade-offs across a complex decision space that includes delivery vehicle size, weight, and power (SWAP), nose cone design, yield, explosives, and logistics trail, to name a few. In both areas, the presence of more resilient and survivable systems and architectures could factor into stability and assurances. Given the reliance on and vulnerability of space systems, current strategic thinking assumes that space assets will be immobilized or neutralized during or prior to nuclear use. Nations use analysis of weapons quantities and capabilities, including accuracy, for operational planning. Increasingly accurate submarine-launched ballistic missiles may improve precision and reduce the number of warheads necessary for effective targeting but could also destabilize dynamics achieved across the nuclear triad and undermine confidence in a country's second-strike capability. Would current assumptions hold if nations develop more survivable space system architectures or more accurate missiles? The strategic effects of AI decision support for system design may not be all negative; modernized systems that are reliable and secure by design may temper anxieties during a crisis or escalation and minimize the risk of accidents during peacetime.

Logistics and Maintenance

Finally, the fourth intersection of AI and the nuclear mission is arguably the most impactful potential application, given the state of technology and current DOD priorities. Underlying nuclear readiness is a complex logistics and maintenance system with the goal of ensuring personnel and equipment perform as expected on a day-to-day basis and in a crisis if called upon. However, the 2014 independent review of the DOD nuclear enterprise found maintenance operations and logistics support pushed to the breaking point, requiring personnel to "make it work" with unreliable test equipment, workdays exceeding 14-16 hour shifts, antiquated and limited facilities, and repair work orders that go unfulfilled for years. While considerable progress has been made to rebuild and strengthen the military's nuclear enterprise, maintenance and logistics remain daunting, especially while awaiting new and modernized systems and facilities. These challenges are exacerbated by the fact that nuclear logistics and maintenance must operate across "a loose federation of separate nuclear activities often imbedded in and indistinguishable from support for and execution of a wide range of non-nuclear activities," contributing to a disconnect in mission ownership throughout chains of command and across services.[40] While the task of gathering, consolidating, and preparing data across the nuclear enterprise is a non-trivial feat for a computing and data infrastructure that is at the initial stages of modernization, it is further complicated by the dispersed nature of operations. Finally, because the nuclear systems are a relatively small portion of the DOD supply chains—systems that are set up to move large quantities of parts relatively often, given the scale and scope—NC3 maintenance and logistics operations often fall low on the priority ladder.

However, commercial sector companies, such as UPS, have demonstrated the potential return on investment to those who invest in analytics and optimization for logistics.[41] The DOD Joint Artificial Intelligence Center (JAIC) successfully deployed ML for preventative maintenance on Blackhawk helicopters in 2020.[42] The JAIC's program used ML to monitor and predict component fatigue and failure patterns in order to replace parts before they failed in service. Similarly applied to the nuclear enterprise's maintenance and logistics activities, ML provides a means to

---

[40] "Independent Review of the Department of Defense Nuclear Enterprise," *TIME*, June 2, 2014, https://time.com/wp-content/uploads/2014/11/external-review.pdf.

[41] Chuck Holland et al., "UPS Optimizes Delivery Routes," *INFORMS Journal on Applied Analytics* 47, no. 1 (January-February 2017), doi:10.1287/inte.2016.0875.

[42] David Vergun, "Delivering AI to Warfighters is Strategic Imperative," U.S. Department of Defense News, September 10, 2020, https://www.defense.gov/Explore/News/Article/Article/2343500/delivering-ai-to-warfighters-is-strategic-imperative/.

discern patterns and failure modes for the wide variety of components necessary to support a reliable nuclear triad and the extensive infrastructure that underpins it.

Recommendations and Conclusion

Existing scholarship at the AI-nuclear nexus is rightfully cautionary on the topic of integrating AI and autonomy into NC3 near a decision to launch nuclear weapons.[43] DOD officials have clearly stated that humans will remain in control of decisions to launch.[44] Alarmingly, perceptions of adversarial willingness to incorporate AI and automation into nuclear systems paint a contrasting picture of international intentions, as seen with Russian development of an autonomous underwater delivery system.[45] However, given the potential consequences, nuclear decisionmaking must be as deliberate and confident as possible about any operational decision to use nuclear weapons. Additionally, injecting AI into the existing NC3 infrastructure does not mean replacing systems that already serve their purpose. Presently, AI and ML may introduce more uncertainty than confidence into such scenarios. It is in these areas of low-probability, high-consequence events and exceedingly short timelines that the immaturity and technical shortcomings of AI are acutely evident.

As legacy NC3 systems age out, the nuclear community risks missing the broader discussion of the merits and drawbacks of incorporating AI and ML into the nuclear enterprise through an almost singularly focused debate on autonomous launch authority. The use of AI and ML in the commercial sector will continue to progress forward, trending toward more robust algorithms that are less brittle and context dependent. Latency, security, and bandwidth considerations for applications of ML are trending toward greater edge computing, where the processing occurs on the device or closer to the source of data.[46] Research is also being directed toward efforts to clearly assess and identify bias in data as well as in algorithmic performance. For national security applications, increasing focus is being paid to the need to establish framework, standards, metrics, and processes for test and evaluation, and verification and validation (TEVV).[47]

The nuclear community should prioritize AI and ML research and analysis in the left of launch functions with the problem types, timelines, and conditions well suited to the technology to better understand the technical risks, escalatory risks, and interplay between the technical and strategic domains. Like all organizations seeking to deploy AI and ML, the nuclear community must not shy away from grappling with questions of responsible use of AI, including data bias, traceability, explainability, predictability, and understandability throughout processes and chains of command. At what point do decisionmakers need to know that AI is informing, in some way, the information they are seeing and the options they may choose from? While this paper focused on the role of AI, a related discussion on the role of automation is also warranted. Modernizations efforts must consider the challenges

---

[43] For example, see: Michael C. Horowitz, "Artificial Intelligence and strategic stability," in *The Impact Of Artificial Intelligence On Strategic Stability And Nuclear Risk Volume I Euro-Atlantic Perspectives*, Vincent Boulanin, ed. (Stockholm, Sweden: SIPRI, 2019), https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf.

[44] Sydney J. Freedberg Jr., "No AI for Nuclear Command & Control: JAIC's Shanahan," Breaking Defense, September 25, 2019, https://breakingdefense.com/2019/09/no-ai-for-nuclear-command-control-jaics-shanahan/.

[45] Valerie Insinna, "Russia's nuclear underwater drone is real and in the Nuclear Posture Review," Defense News, January 12, 2018, https://www.defensenews.com/space/2018/01/12/russias-nuclear-underwater-drone-is-real-and-in-the-nuclear-posture-review/.

[46] Paul Miller, "What is edge computing?," The Verge, May 7, 2018, https://www.theverge.com/circuitbreaker/2018/5/7/17327584/edge-computing-cloud-google-microsoft-apple-amazon.

[47] Michèle A. Flournoy, Avril Haines, and Gabrielle Chefitz, "Building Trust Through Testing: Adapting DOD's Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, Including Deep Learning Systems," WestExec Advisors, October 2020, https://cset.georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf.

of integrating new technology alongside legacy systems. Finally, additional scholarship is critically needed to bring together the AI and nuclear communities to better understand the nature of AI capability, calibrated trust in ML for nuclear applications, stability inflection points, and strategic effects.

AUTHOR

Lindsey Sheppard *is a fellow in the International Security Program at the Center for Strategic and International Studies (CSIS), where she focuses on the nexus of emerging technologies and national security for the United States and allied and partner nations. Her research areas include artificial intelligence, machine learning, autonomous systems, defense innovation policy, and technology ecosystems.*

ABOUT CSIS