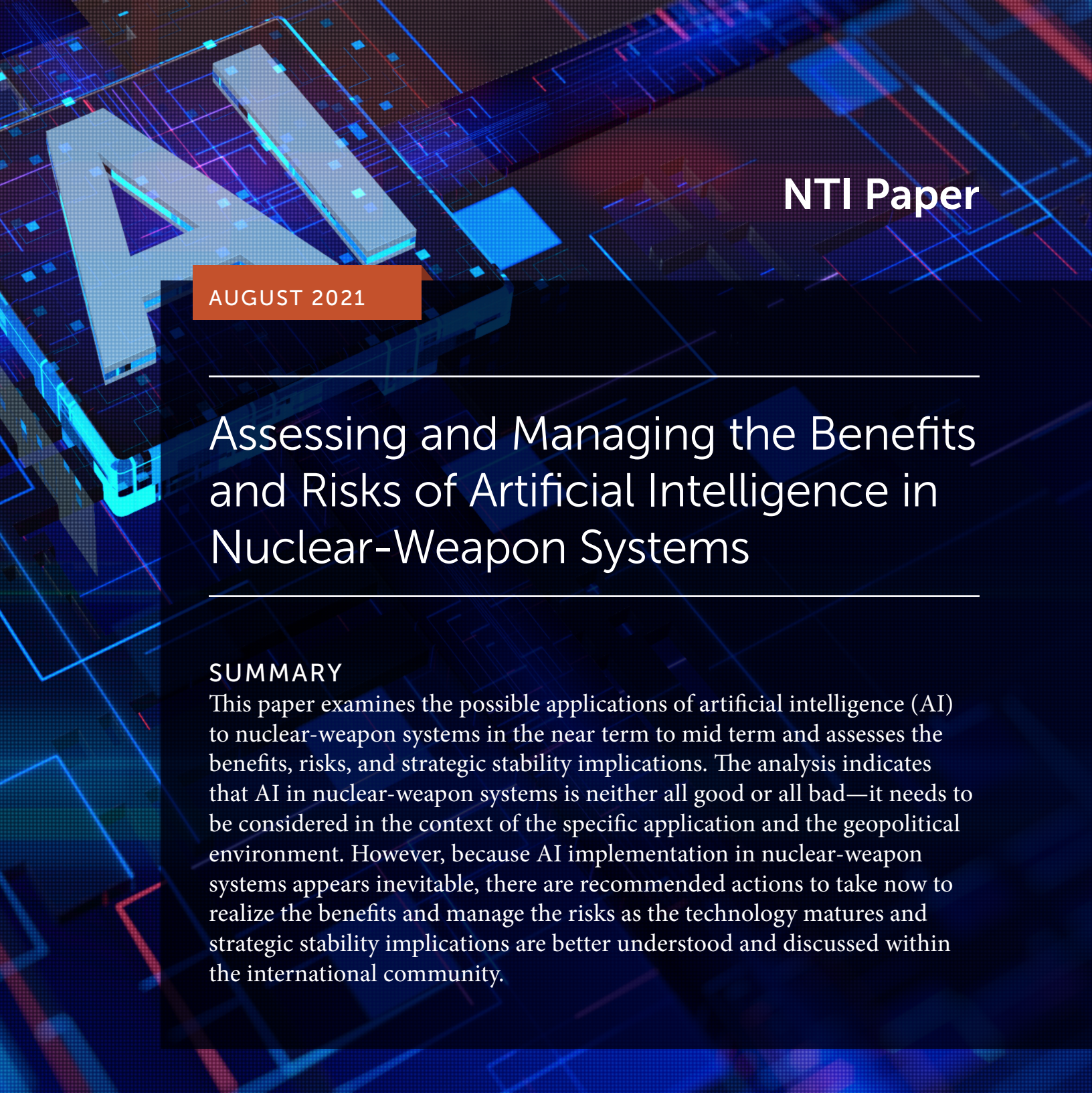


NTI Paper

AUGUST 2021

Assessing and Managing the Benefits and Risks of Artificial Intelligence in Nuclear-Weapon Systems

SUMMARY

This paper examines the possible applications of artificial intelligence (AI) to nuclear-weapon systems in the near term to mid term and assesses the benefits, risks, and strategic stability implications. The analysis indicates that AI in nuclear-weapon systems is neither all good or all bad—it needs to be considered in the context of the specific application and the geopolitical environment. However, because AI implementation in nuclear-weapon systems appears inevitable, there are recommended actions to take now to realize the benefits and manage the risks as the technology matures and strategic stability implications are better understood and discussed within the international community.

Jill Hruby and M. Nina Miller

20 YEARS OF NTI BUILDING A SAFER WORLD

Contents

Executive Summary

At a time when most countries with nuclear weapons are modernizing or diversifying their nuclear arsenals, significant technological advances in artificial intelligence (AI) for military applications suggest that AI inevitably will be explored for use in nuclear-weapon systems. Along with significant benefits, however, come associated risks and implications for strategic stability.

At a time when most countries with nuclear weapons are modernizing or diversifying their nuclear arsenals, significant technological advances in artificial intelligence (AI) for military applications suggest that AI inevitably will be explored for use in nuclear-weapon systems. Along with significant benefits, however, come associated risks and implications for strategic stability.

Two application areas are considered the most likely to take advantage of AI advances in the near term to mid term: Nuclear Command, Control, and Communications (NC3) and autonomous nuclear-weapon systems. This paper envisions the specific functions AI could perform in these two areas and analyzes the potential positive and negative consequences.

In NC3, AI could be applied to enhance reliable communication and early warning systems, to supplement decision support, or to enable automated retaliatory launch. The implications vary dramatically. Enhancing communication reliability and decision-support tools with AI has recognizable benefits, is relatively low risk, and is likely stabilizing, although it still requires additional technical research to lower risk as well as deeper policy exploration of stability implications to avoid provoking an arms race. AI application to automated retaliatory launch, however, is highly risky and should be avoided.

For autonomous nuclear-weapon systems, AI along with sensors and other technologies are required for sophisticated capabilities, such as obstacle detection and maneuverability, automated target identification, and longer-range and loitering capability. Today's technology and algorithms face challenges in reliably identifying objects, responding in real time, planning and controlling routes in the absence of GPS, and defending against cyberattacks. Given the lack of technology maturity, fully autonomous nuclear-weapon systems are highly risky. These risks, combined with the potential instability these weapons may cause, suggest that a ban on fully autonomous systems is warranted until the technology is better understood and proven.

For each state with nuclear weapons, the specific application and timing of AI incorporation will depend on the production or modernization schedule, the perceived benefits and needs, the technical capabilities and level of investment, and the level of risk acceptance. To encourage safe application and help minimize risks and negative effects on strategic stability as AI is introduced into nuclear-weapon systems over time, the following is recommended:

- The U.S. national security enterprise should prioritize research on low technical risk approaches and fail-safe protocols for AI use in high-consequence applications. The research should be openly published as long as it does not jeopardize national security. Additionally, cooperative research

with international partners should be considered, and other states with nuclear weapons should be encouraged to conduct research with the same purpose.

- States with nuclear weapons should adopt policies and make declaratory statements about the role of human operators in nuclear-weapon systems and/or the prohibition or limits of AI use in their nuclear-weapon systems.
- The international community should increase dialogue on the implications of AI use in nuclear-weapon systems, including how AI could affect strategic and crisis stability, and explore areas where international cooperation or development of international norms, standards, limitations, or bans could be beneficial.
- In addition to an analysis and recommendations, this paper offers a summary of AI technology relevant to nuclear-weapon systems to provide background for those not already well versed in AI.

Introduction

Artificial intelligence (AI) is widely considered to be a catalyst for revolutionary change, with application in fields as diverse as finance, transportation, health care, and war fighting.[1] Within those fields and others, the speed of technological development is generating both enthusiasm for AI's benefits and alarm over the potential for doomsday scenarios.[2] The fast-paced development and implementation of new AI technologies also is compelling the establishment of national AI ethical standards and policies in an attempt to realize the benefits of the advancing capability while discouraging negative societal consequences.[3] The military application of AI is generating particular attention, and some countries have initiated open discussion around policies governing this area of exploration. At the same time, the United Nations has held talks concerning lethal autonomous weapon systems on the battlefield, and approximately 30 state parties have called for a treaty banning those systems.[4]

Today's military applications of AI include intelligence collection and analysis, logistics, cyber operations, information operations, command and control, and autonomous or semi-autonomous vehicles.[5] Many of these AI military applications aim to increase the efficiency of tasks that are dull, dirty, or dangerous.[6] Reconnaissance drones and battlefield management systems that employ AI already are in use with a level of international acceptance and adoption. As AI advances, its military use could revolutionize war fighting with the realization of concepts such as autonomous vehicle swarming.

Today's military applications of AI include intelligence collection and analysis, logistics, cyber operations, information operations, command and control, and autonomous or semi-autonomous vehicles.

Because of vast application areas for AI, it is not surprising that the United States, Russia, and China each view the military potential of AI as highly important, although the specific priorities are somewhat distinct. The U.S. military looks at AI to maintain general technological superiority. Russian President Putin has said whoever leads AI will "become the ruler of the world,"[7] and he has advocated Russian military use of AI especially for autonomous robots and vehicles. The Chinese military has pursued the opportunities AI offers to overtake American advantages and has a strategy to outpace the United States through widespread "intelligentized" warfare and civilian-military fusion.[8] China's military interests are heavily focused on machines augmenting human decision making and performance, as well as swarming autonomous systems. The Russian and U.S. militaries have actively experimented with AI in regional military engagements, whereas China has not used AI on the battlefield.[9] A United Nations workshop held in 2019, "The Militarization of AI," concluded that the implications for international peace and security of AI military efforts remain unclear, and that the uncertainty generates fear and heightens perceptions of risk.[10]

AI technology is relevant to military and commercial applications, including conventional and nuclear weapons. Because of the high potential consequences, AI use in nuclear-weapon systems seems to be proceeding at a slower pace—or perhaps more covertly—than other military applications. Nonetheless, the

potential for AI application to nuclear-weapon systems is likely to grow as the military use of AI develops concurrently with nuclear-weapon systems modernization and diversification.[11]

The advancement and maturing of AI in commercial and conventional weapons applications offers insight into potential applications in nuclear-weapon systems. Careful examination of the specific potential nuclear-weapon systems applications is needed to understand the benefits, risks, and stability implications to guide both technical and policy considerations going forward as well as to initiate meaningful international dialogue. Several studies have been published outlining the potential applications of AI in nuclear-weapon systems,[12] and the Nuclear Threat Initiative (NTI) has published a paper on the security and policy implications of integrating digital technology, including AI, into U.S. nuclear-weapon systems.[13] This paper aims to deepen the discussion by detailing potential applications along with their benefits, risks, and strategic stability implications. It offers recommendations for the technical and policy communities to consider before AI applications to nuclear-weapon systems are further developed and implemented.

The paper is organized in three parts:

- **"Artificial Intelligence Characteristics Relevant to Nuclear-Weapon System Applications"** describes the characteristics that could provide the most benefit or introduce the most risk.
- **"AI Application to Nuclear Weapons and Their Operational Systems"** addresses the potential AI application to nuclear-weapon systems and is the core of the paper. The focus is on applications that can be imagined in the near term (0–5 years) and the mid term (5–10 years), specifically Nuclear Command, Control, and Communications (NC3) and autonomy. There are other potential AI applications in areas relevant to the supporting nuclear enterprise, including war gaming and planning, physical security, missile defense, and nuclear proliferation and arms control agreement monitoring. These are briefly discussed but are not the focus of this paper.
- **"Recommendations"** recognizes the expected use of AI in nuclear-weapon systems as well as the desire to keep nuclear risks low and sustain or enhance nuclear strategic stability.

Artificial Intelligence Characteristics Relevant to Nuclear-Weapon Systems Applications

The following basic description of AI, including the benefits and risks that could be applicable to nuclear-weapon systems, is provided as context for the remainder of the paper. This paper does not review AI comprehensively because it is widely documented and rapidly advancing, and not all of it is relevant to understanding how AI might be used in nuclear-weapon applications.[14]

Rules-Based AI and Machine Learning

The term "artificial intelligence" colloquially refers to computational processes that perform functions usually completed by people. Today's AI systems are tools and methods that allow computers to execute rules-based commands (first-wave AI) or use data and machine learning (ML) techniques to categorize information (second-wave AI). First- and second-wave AI is sometimes referred to as "narrow AI." The Defense Advanced Research Projects Agency discusses the third wave of AI as contextual adaptation,[15] and others often refer to this as "artificial general intelligence" (AGI). Contextual adaptation or AGI is widely believed to be a decade or two away from maturity.[16] However, today's narrow AI capabilities are growing in sophistication and are increasingly widely deployed.[17]

The basic elements of AI today are shown in Figure 1 (courtesy of MIT Lincoln Lab).[18] AI depends on data from sensors or other sources as well as simulated data. Because data can be collected from multiple

Figure 1: AI System Architecture

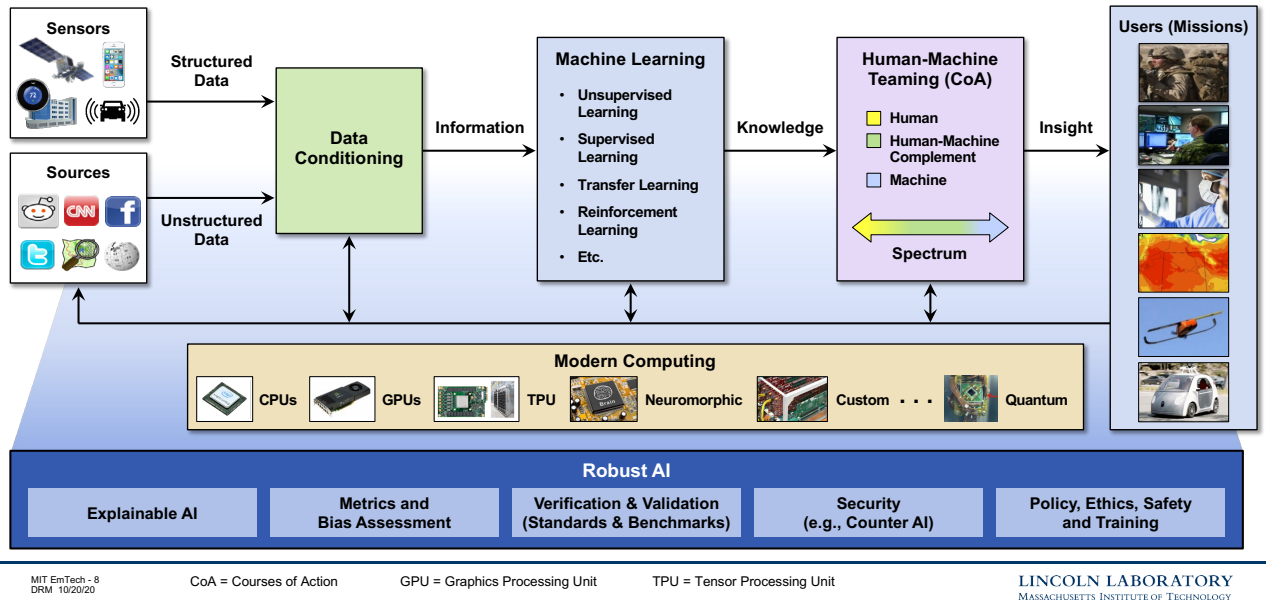


Diagram courtesy of David Martinez, MIT Lincoln Laboratory.

sources, merging the data into common formats, or conditioning the data, often is needed. Once the data are conditioned, the information is used by various types of algorithms to create knowledge for use by either people, machines, or teams composed of both. Once people and/or machines evaluate the knowledge, the insight gained can be applied to a particular objective. When people are not involved in evaluating the data, the system is considered autonomous. Fully autonomous systems that use AI are still rare. The AI process depends on and is enabled by modern computing, raw data, and data fed back from application.

Rules-based AI uses human-crafted "if-then" structured reasoning to enable computers to determine the appropriate responses to input. Rules-based AI has been broadly used but has limitations due to the intense human labor needed to create the rules. Examples of rules-based AI to high-consequence applications include feedback control systems for nuclear power plant operations and aircraft autopilot systems. Experts Michael Horowitz, Paul Scharre, and Alexander Velez-Green asserted in an article for Cornell University that AI for nuclear systems will likely be rules-based AI because of the high consequence of failure.[19]

Because of its ability to learn through data, ML is particularly useful where the relationship between the data and the result cannot be described analytically.

Today, the term AI most often refers to machine learning or the fast-growing subfield of neural networks.[20] ML identifies patterns in relatively large data sets through inference, and results are fundamentally probabilistic.[21] Everyday ML applications include facial and voice recognition as well as predictive suggestions for books or movies people may like based on other purchases or reviews. Self-driving cars are perhaps the best-known application that requires sophisticated ML (combined with other automotive and technological advances).

This paper imagines artificial intelligence based on either rules-based or ML approaches that could be used in nuclear-weapon systems.

The Potential Benefits of AI Systems

Rules-based AI systems perform efficiently and reliably for a well-defined set of conditions. These first-wave AI systems often can outperform human operators by being able to function consistently and continuously and perhaps also by processing information more quickly. Rules-based AI can be used effectively to augment human performance and speed in applications that require continuous monitoring.

Because of its ability to learn through data, ML is particularly useful where the relationship between the data and the result cannot be described analytically. As a result, ML can produce findings that are not necessarily causal and can correlate two unrelated variables with a third unknown variable without a rational causal relationship. Although this can generate unsafe suggestions, it also can produce novel or more efficient approaches.[22] ML algorithms can therefore make predictions that may be accurate while appearing illogical to people. A well-known example of the power of ML is the success of Alpha Go, the DeepMind program that beat human players at the board game Go using moves never observed in human play.[23] In the security arena, ML can be used to sort through hours of reconnaissance video images to find abnormal patterns that should be further examined by a person.

The Potential Risks of AI Systems

There are considerable well-documented risks involved in AI systems.[24] The risks deemed the most relevant for nuclear-weapon systems are discussed briefly here.

Data Insufficiency

Because AI systems depend on data, when robust data are not available, the systems can fail catastrophically—sometimes referred to as "brittle failure." In rules-based systems, failure occurs when a condition is presented for which there is no rule provided. In ML systems, brittleness occurs when the system encounters a situation or context for which it has not been trained. ML typically covers many more conditions than rules-based systems if there are sufficient training data.

When an AI system will be operating in an environment that lacks comprehensive data sets, simulated data may be added to actual data. The limitation of real-world data sets in the case of nuclear-weapons applications necessitates that some data will need to be simulated. For example, although data exist for the signatures associated with the launch of a ballistic missile, they may need to be augmented with simulated data based on physical modeling of rocket combustion for newer or less tested missiles. However, in other cases, such as failure of an early warning system, there are (fortunately) little data, and simulated data would need to be constructed carefully to ensure completeness and accuracy.

Data Corruption

ML-based AI is particularly vulnerable to data poisoning and adversarial data spoofing. Data poisoning attacks involve introducing training data that cause a learning system to make mistakes. Such manipulation is often difficult to detect until after a mistake is made. Adversarial data spoofing, or inputs designed to cause a model to make a mistake, can manipulate an AI algorithm to generate flawed predictions.[25] Such spoofing often is demonstrated by applying an imperceptible non-random perturbation to an image that causes the AI/ML algorithm to wrongly categorize it.[26]

Attacks that use poisoning, misdirecting, and disabling an opponent's systems are a known commodity of intelligence and military operations.[27] Before AI is deployed in a nuclear-weapon system, the algorithms should be rigorously tested against data poisoning or spoofing, followed by continuous validation and verification of the deployed systems, to ensure that an attack would be promptly identified. Test, evaluation, validation, and verification protocols for AI systems are in the very early stages of development and need to improve before safe application in military systems, especially those that authorize force.[28]

Inability to Explain or Understand Results

AI systems that use ML can produce results that are not intuitive and cannot be understood through simple logic. For example, the mathematical parameters determined by a neural network analysis can be determined, but the reason those parameters were derived may be impossible to establish. There is active research, often called "explainable AI," to better understand the underlying logic of ML systems, but this lack of transparency remains a serious flaw.[29] The inability to understand why algorithms produce specific results is cause for concern, especially if decision makers are using the results in high-consequence applications.

Bias

At least two types of bias exist in AI systems. One type of bias is associated with human decision makers placing too little or too much trust in AI results, referred to as either a "trust gap" or "automation bias," respectively. The other type of bias is associated with inherent bias in data sets or rulemaking that is transferred from human behavior, likely unwittingly.

Human decision makers have a complex relationship with machines[30], and researchers have demonstrated an early tendency for people not to trust AI recommendations—especially if it goes against their experience or intuition. This is often referred to as the trust gap. Some research suggests that over time, the opposite tends to occur, with people trusting the results of automated decision making even when the information goes against their expert judgment. This is often referred to as automation bias. Although still an area of active research, it is clear that human-machine partnerships are complex and that any bias—whether it is overly skeptical or overly trusting—is a risk when incorporating AI, especially in an advisory capacity.[31]

Although still an area of active research, it is clear that human-machine partnerships are complex and that any bias—whether it is overly skeptical or overly trusting—is a risk when incorporating AI, especially in an advisory capacity.

Inherent bias has also been found in several AI applications. For example, Amazon's AI hiring assistant was rejecting female applicants because most previous hires were male.[32] In the case of nuclear-weapon systems, inherent bias or misunderstanding about an adversary or its response could be inadvertently encoded in rules or data.

Lack of Open-Source Community

The open-source community has been essential for ML development and success. Open-source communities, supported by companies including Microsoft, Alibaba, Google, and Facebook, demonstrate that data are perhaps more important than the models themselves.[33]

When privacy or security concerns exist, this open-source system is not as viable. ML success in the private sector's open-source environment may not translate to military applications. During the development and acquisition process of an AI system, a key concern will be understanding what feasibility is lost if feedback loops are secret and/or slow. Whereas open-source researchers benefit from compiling and sharing large data sets, AI in the defense community would need training data from specialized or perhaps classified environments. Military investment in AI may struggle to develop private or classified sharing models. The *Artificial Intelligence Strategy* from the U.S. Department of Defense (DoD) recognizes this challenge and articulates initiatives to address it, including forming open-mission initiatives focused on global challenges such as disaster relief; strengthening partnerships with academia, U.S. industry, and global partners; and engaging the open-source community.[34]

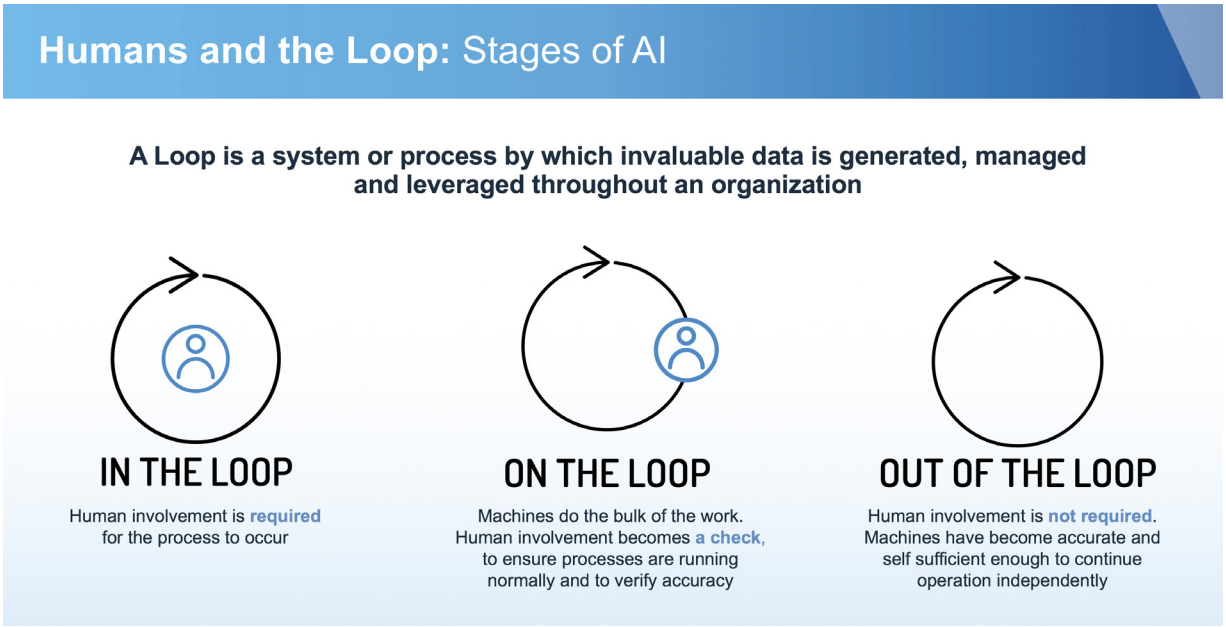
Managing AI Risks through Human Involvement

In addition to well-tested, validated, and verified systems, another approach to manage AI risks is to ensure a human is always involved. As shown in Figure 2,[35] there are three types of human supervision: a human can be *in-the-loop*, meaning a human will make final decisions; the human can be *on-the-loop*, supervising the system and data being generated; or the human can be *out-of-the loop* for full autonomy.[36] The proper integration of people and machines requires considerable care, and this is another area of ongoing research.

The commitment from the United States, Russia, and China to maintain human engagement in military and nuclear application varies in transparency and continues to evolve:

- In the United States, a 2012 DoD directive established that lethal autonomous systems must "allow commanders and operators to exercise appropriate levels of human judgment over the use of force."[37] Although no public statement exists regarding automated control of nuclear-weapon systems, analysts agree the United States will maintain human-in-the-loop authority for AI/ML systems.[38] Congressional testimony in February 2020 by Gen. Terrence J. O'Shaughnessy, then-commander of NORTHCOM, advocated for human-on-the-loop instead of human-in-the-loop for nuclear missile *detection* to realize the increased speed of machines compared to humans.[39] All current statements suggest that the United States will maintain human decision making for the use of nuclear weapons for the foreseeable future, even if that person may be receiving decision support from an AI-enabled system.
- A Russian non-proliferation researcher, Petr Topychkanov, says that "autonomy has been widely used in [the Soviet Union and Russia's] nuclear command-and-control, ballistic missile defence, early-warning and now strike capabilities. But it has never replaced the human in the loop."[40] It is

Figure 2: Roles for Humans and Machines in Decision Making



Source: www.datacenterdynamics.com/en/opinions/path-ai-connected-government.

widely believed there is human-in-the-loop authorization required before Perimeter, the Russian semi-automated retaliation system, would launch nuclear weapons.[41]

- China has not publicly asserted a policy for human supervision of AI in nuclear-weapon systems.[42] The Beijing AI Principles developed and endorsed by researchers and industry experts, published in 2019, do not explicitly refer to the deployment of AI in nuclear weapon or military systems.[43] However, China has suggested in general terms that it would examine the legal, ethical, and moral implications of AI before integrating it into new political or military tools. The lack of clear articulation regarding human involvement in nuclear command and control is nonetheless concerning.

Relying on people to correct mistakes made by machines is not necessarily sufficient to catch or avoid errors due to a trust gap or automation bias. However, human-machine partnerships are seen as more reliable than humans being out of the loop.[44] Machine-plus-human decision-making systems still require considerable research to understand, optimize, and implement.

AI Application to Nuclear Weapons and Their Operational Systems

There are two broad areas where AI will likely intersect with nuclear weapons and their operational systems in the near term to mid term:

- Nuclear Command, Control, and Communications
- Autonomous nuclear-weapon systems

These potential application areas are broken into their individual components, and this section describes the associated benefits, risks, and implications for strategic stability.

The degree to which any of these applications of AI are being used by specific countries is evolving and often not transparent. The United States, Russia, and China all have established a high priority for AI research and use, and there is clear indication that AI is being explored for military as well as commercial application in these countries and others.[45] Yet the U.S. DoD Joint Artificial Intelligence Center does not include any nuclear-related activities in its published prioritized areas of research.[46] There is little open-source description of AI activities aligned with nuclear-weapon systems in either Russia or China, although both have announced the development of new autonomous (although perhaps actually semi-autonomous) nuclear-weapon or dual-use delivery systems. Included below is information found concerning the direct or indirect use of AI in nuclear-weapon systems. However, most of the applications described here are hypothesized on the basis of knowledge of AI and nuclear systems and are included to consider the possible future uses of AI and their implications.

Nuclear Command, Control, and Communications

The potential intersection of AI with NC3 is significant and perhaps the most readily imagined and feared intersection, due to the high consequence of automated retaliatory launch of nuclear weapons, as portrayed in movies such as *Dr. Strangelove* and *Wargames*. For those in the nuclear field, discussion of AI in NC3 recalls the long-ambiguous Soviet system called Perimeter (alternatively spelled Perimetr), often referred to in the West as "The Dead Hand," the title of journalist David Hoffman's book[47] on the Cold War arms race.

There is considerable contemporary interest in the intersection of AI with NC3 because the United States has committed to modernizing its NC3 system to account for current and future threats, and to synchronize information across nuclear and non-nuclear command and control.[48] The current U.S. NC3 system is more than 30 years old, is made up of more than 200 individual systems,[49] and was designed to counter a massive nuclear attack from Russia. The system has been updated with more modern technology and modified over the years to address changing adversarial conditions. In 2018, a focused modernization effort of the U.S. NC3 enterprise was started,[50] and the U.S. 2018 Nuclear Posture Review[51] discusses improvements needed to NC3 systems, including the following statement: "The United States will continue to adapt new technologies for information display and data analysis to improve support for Presidential decision making and senior leadership consultations." This statement is not explicit regarding what technology or procedures may be

deployed, but it suggests that new visual decision aids are needed and potentially hints at the potential use of AI as a data-analysis or decision-support tool.

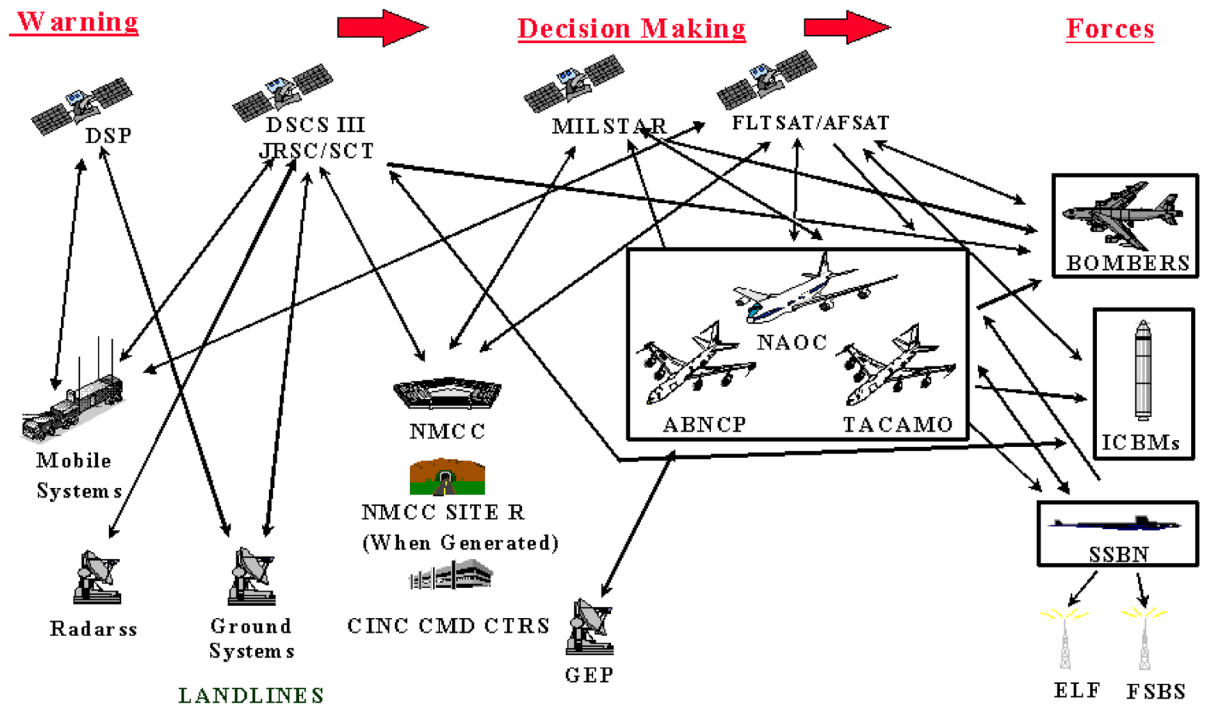
There are initial indications that China might incorporate AI/ML into nuclear-relevant systems, including for improved early warning, targeting, and command and control.[52] Russia's leadership appears to be pursuing a strategy involving increased reliance on AI and automation in nuclear and conventional command and control.[53]

Four areas where AI could be used in NC3 are considered here: (1) reliable communication, (2) early warning systems, (3) decision support, and (4) automated retaliatory launch. Some discussions of AI by other authors separate early warning from command and control, but because they are a continuum of capabilities and usually grouped together by the U.S. DoD, they are discussed here in the same section. Any incorporation of AI in NC3 systems comes with benefits, risks, and implications to strategic stability.

Reliable Communication

Reliable communication is a cross-cutting element of NC3. Early warning, decision conferencing, receiving orders, and managing and directing forces all require reliable communication links. For NC3 communication systems to be reliable, they must remain secure and be resilient to physical attack or cyberattack. NC3 communication systems are complex with many nodes.[54] As an example of the complexity, the notional connectivity of the U.S. NC3 system is shown in Figure 3.[55] AI has the potential to improve the security

Figure 3: Notional Connectivity for Nuclear Command and Control in the United States



Source: "National Command & Control: The National Military Command System (NMCS)," October 2001.

of the communication system if used to defend against cyberattacks. It also has the potential to improve the resilience of the communication system if used for real-time health assessment and optimization of communication pathways.

Cybersecurity

AI could be a necessary tool to improve the cybersecurity of a modernized, digital NC3 system—although, ironically, any AI system introduces a new vector for a cyberattack, and AI could also be used in the future as an adversarial tool to automate cyberattacks. Both the benefits and the risks of using AI in NC3 cybersecurity are important to consider and anticipate.

It is widely recognized that the potential for a cyberattack on an NC3 system is significant because it is a critical node. Even systems that are believed to be digitally unconnected to the external world, such as Perimeter or the current U.S. NC3 system, could have vulnerabilities that could lead to an external penetration and corruption of data or algorithms.[56] In a modernized NC3 system, there could be many digital systems, including those used to collect and transmit data and images, to allow secure communication between locations of military and political decision makers, and to transmit command orders.

Any AI system must be robust to cyberattack or obfuscation, and if AI is adopted too quickly and without cybersecurity standards, it could increase cyber risks.

Automated AI-enabled cybersecurity systems have been in use since at least 2005 in other applications.[57] Examples of such AI-enabled, automated cyber defense systems include spam filters, antivirus engines, heuristic intrusion-detection mechanisms, endpoint detection, and response solutions.[58] As AI advances, AI techniques could further improve automated detection efficiency by, for example, deploying multiple algorithms simultaneously, and they could help in attribution by identifying patterns of behavior. Although implementation of automated cybersecurity in nuclear systems may be proceeding at a slower pace because of high security standards and older digital systems, using AI in NC3 systems to automate and enhance cyber threat detection and response seems likely and maybe even necessary in a modernized NC3 system with digital capabilities. However, even AI-enabled defense is not guaranteed to detect all attacks, especially sophisticated techniques such as Zero-Day exploits, unless new breakthrough approaches are developed.[59]

Unfortunately, once AI is introduced into a cybersecurity system or any other system, there are new pathways to exploit the system by subverting either the AI algorithms or the incoming data stream.[60] Therefore, the AI-enabled cyber-detection system design must be especially robust and constantly validated and verified to detect any adversarial attempts to subvert the algorithms. In addition to tampering with algorithms, adversaries could attempt to intercept and subvert images or signals to, for example, hide a missile launch. A third party could potentially exploit AI in an NC3 system by sending fake signals or images of an incoming missile attack as a diversionary tactic. Detecting and remedying cyberattacks on ML algorithms is made more difficult by the underlying opacity of the machine's decision process.[61] Any AI system must be robust to cyberattack or obfuscation, and if AI is adopted too quickly and without cybersecurity standards, it could increase cyber risks.[62]

Of significant concern going forward is the use of AI to automate cyberattacks and to power learning for adaptive attacks.[63] An AI-enabled cyberattack could, for example, increase the stealth of attacks, employing channels with similar web addresses to that of an infected organization or taking advantage of popular ports.[64] Attackers today are trying to abuse the AI systems used by defenders, but experts believe they are not yet successfully using AI to learn and improve attack vectors.[65]

Real-Time Optimization of Communication Pathways

In addition to using traditional techniques to harden communication networks, AI could play an important role in improving the resilience and reliability of NC3 communication systems. Today, a common method for resilience is to make sure that space, air, and terrestrial communication pathways exist for redundancy and potential use. In the future, AI in combination with continuous network monitoring could be used to identify the optimal communication pathway at any instant in time, including by establishing new routes as needed. This type of resilience and reliability in communication for NC3 is especially important in crisis scenarios.

AI use in communications is an expanding area of research,[66] and it is likely such work will generate other ways to improve communication privacy and security that also could assist in operating through contested communications environments.

Benefits, Risks, and Strategic Stability Implications of AI for Reliable Communication

Implementing AI to improve reliable communication by enhancing cybersecurity and optimizing communication pathways in modernized NC3 systems has potential benefits to operational performance, as described above.

Nonetheless, additional research on managing the risk of any embedded AI system is needed. For example, robust validation and verification methods should be established before AI incorporation in high-risk systems. Because there is an opportunity to learn from the early application of AI for cybersecurity and reliable communications in non-nuclear systems, it seems prudent to observe the vulnerabilities introduced and optimize test protocols in lower-risk military applications before any application to NC3.

In general, improved cybersecurity enhances strategic stability because a country with nuclear weapons can be reassured the system will work as planned, and any adversary will be less confident that hacking would succeed and therefore less likely to depend on the development and deployment of cyber as a weapon. However, strategic instability could result if AI is used effectively by one state with nuclear weapons to increase cybersecurity or challenge AI-enabled cyber defenses and not by others. Because many of today's states are sophisticated cyber actors, it is likely some will deploy sophisticated cybersecurity. Therefore, in the near term to mid term, AI-enabled cybersecurity incorporation is likely by at least some states with nuclear weapons—and is potentially stabilizing. However, this is an area where perceptions are as important as ground truth and where continued competition will occur unless international norms of behaviors are adopted and adhered to by all states with nuclear weapons.

AI for real-time communication optimization in NC3 also is likely to be beneficial to strategic stability because it helps the system operate as intended even under attack. Communication systems can be damaged by direct attack from either nuclear or precision conventional weapons, can be damaged from

cyberattack, or may be unintentionally damaged as collateral.[67] Over the past decade or so, the probability of a communication system attack has increased in part because of nuclear and conventional communication systems entanglement. Additionally, current U.S. and Russian[68] nuclear doctrines state that attacks on critical infrastructure could trigger a nuclear response. However, if a communication system could quickly recover from an attack by automatic rerouting, perhaps the probability of nuclear escalation from an NC3 attack would decrease because the damage could be managed. For these reasons, AI use to enable reliable communication even under attack provides a buffer against quick escalation and is likely stabilizing.

Early Warning Systems

One of the five critical functions of an NC3 system is the string of three tasks sometimes referred to as early warning: detection, warning, and attack characterization.[69] In the United States, early warning is alternatively referred to as integrated tactical warning and attack assessment.

Over the past decade or so, the probability of a communication system attack has increased in part because of nuclear and conventional communication systems entanglement.

Early warning systems in the United States and Russia have been designed to detect ballistic missiles by identifying launch with space-based infrared (IR) sensors and corroborating flight trajectory with sophisticated ground-based radars. China currently relies entirely on ground-based sensors,[70] but in 2019, Russia announced it was assisting China in developing an early warning system that is thought to include space-based detectors.[71] In space- and ground-based dual-phenomenology systems, the faster the IR and radar signals are detected, analyzed, and compared, the faster attack characterization can be completed, and the more time remains for leaders to decide how to respond. However, speed cannot sacrifice accuracy, as it is essential that detection, warning, or attack characterization does not result in a launch under attack decision using mistaken information.

In the United States, satellite detection of a possible missile launch triggers an alert at the North American Aerospace Defense (NORAD) Command in Colorado. There, military analysts work to confirm that the satellite detection is legitimate and corroborated by ground-based radar, and they validate the calculations of the trajectory and expected target of the missile.[72] If analysts believe the threat is authentic, the information is quickly submitted to military and political leadership. If the threat is deemed erroneous, notification is truncated.

The data being relayed to NORAD are becoming increasingly sophisticated. For example, in the ongoing Future Operationally Resilient Ground Evolution Mission Data Processing Application Framework program, the missile warning satellite data collected are being merged and analyzed along with data from other government satellites to obtain more complete awareness.[73] Similarly, U.S. Upgraded Early Warning Radars are required to be capable of detecting and tracking multiple targets that would be indicative of a massive missile attack. The system must rapidly discriminate between vehicle types, calculate their launch and impact points, and perform scheduling, data processing, and communications. The operation is semi-automatic and relies on highly trained personnel for monitoring, maintenance, prioritization, and scheduling and to serve as a final check of validity.[74] The processed data from both of these systems are relayed to NORAD, where operators generate a common operating picture, potentially already with the help of AI.[75]

In Russia, when the detection system records a missile launch alert, an automated warning is transmitted to leadership. By a separate channel, a human operator on duty reports that an automatic warning was issued and can submit an argument that the information should be blocked from further transmission due to a suspected error.[76] The details of the Russian data-processing capabilities are not known; however, a video released in February 2021 for the 50th anniversary of Russia's early warning system showed an automated trajectory of a ballistic missile launched from the United States projected on a digital map.[77] China's early warning detection system and notification process is not publicly disclosed, although studies indicate that there are concerns about missing an attack (false negative) and therefore a desire to improve the national early warning systems.[78]

Hypersonic systems add to the complexity of early warning because the current ground-based radar systems will detect them later in their flight, and their maneuverability confounds target identification.

Complicating early warning over the past decade is the accelerated development of hypersonic nuclear delivery systems.[79] Hypersonic systems add to the complexity of early warning because the current ground-based radar systems will detect them later in their flight, and their maneuverability confounds target identification. Because hypersonic systems have high temperatures when gliding, some believe they could be tracked in flight by existing IR satellites.[80] However, it is widely believed that more space-based sensors are needed for tracking their full flight trajectory. Some suggest a network of 24 satellites could ensure that missiles launched from anywhere on earth would be within view of at least two sensors for the entirety of the midcourse phase of their flights.[81] The U.S. DoD is now contemplating construction of a space sensor layer, a constellation that would link dozens to hundreds of satellites in multiple orbits to track hypersonic weapons while simultaneously increasing resiliency from attacks on space-based assets. A space sensor layer will require both increased sensor capabilities and AI for data analysis, as well as low latency communications to process data and provide battle management recommendations to commanders.[82]

AI has the potential to allow today's early warning systems designed for ballistic missiles to be faster and more capable and may already be under development or in limited use for some early warning functions. If new types and/or larger numbers of sensor systems are introduced, then AI combined with the new sensors likely increases the probability of AI being used in early warning systems.

Missile Detection and Warning

In the simplest scenario, an AI system could potentially enhance the speed and completeness of the analysis done at the satellite and radar data sites before submission to NORAD. Additionally, AI could be used at NORAD to fuse the two incoming data streams with other Intelligence, Surveillance, and Reconnaissance (ISR) information to add to the confidence of the warning.[83] It has been reported that NORAD already is partnering with the Defense Innovation Unit to create an AI system called Pathfinder to fuse data from military, commercial, and government sensors to help with early warning.[84]

However, the potential for AI to increase the speed, robustness, and ability of detection and warning increases more significantly when AI is coupled to advanced sensor networks: ground, air, space, sea, or undersea based. The deployment of increased numbers and types of sensors is likely over time not only because new nuclear hypersonic missiles need to be tracked but also because sensors are getting cheaper

and more capable and because larger sensor networks themselves are considered more resilient. With additional sensor data, an AI system may be needed because it would be difficult for humans to process the information in a timely manner.

If enhanced sensor systems are deployed, an AI system could, for example, incorporate all space-collected information and make predictions, even before a ground-based radar signal is detected, to increase the speed of warning and attack characterization. For hypersonic missiles, it also is possible that only space-based assets would be used. Such a system might still rely on dual phenomenology but by using different types of space-based sensors instead of IR space and radar ground sensing.

Similarly, AI could improve detection robustness by, for example, comparing the IR signature detected with IR signatures from other known missile launches at the location detected to validate a credible launch. Also, AI-augmented ISR could combine satellite sensor data, other sensor data, and intelligence data to prove, question, or refute the probability of launch and potential trajectory or target.

With an increased number or types of sensors and AI analytics, the flight path for hypersonic missiles could be tracked in close-to-real time, although the desire is to both track and predict the path and target as quickly as possible. To predict hypersonic missile trajectories and likely targets, an AI system trained on possible missile types and their associated flight characteristics could match the possible trajectories, including potential manuevers, to probable high-value targets. For hypersonic nuclear-capable missiles, there is little data available for use in machine learning. In the absence of data collection from large test programs, design information and accurate models of hypersonic missile flight would be needed to create synthetic data. Even if synthetic data can be created, it is still no substitute for access to large amounts of real-world data to build robust and successful algorithms.[85] Once trained, the AI system could use an observed flight to anticipate possible remaining flight trajectories and potential targets. The longer that flight data are obtained, the better chance an AI system will have for accurate prediction. However, the longer the system takes to predict a target, the less time there is for a decision maker to consider a response. The proper balance would require careful consideration.

Considerable effort is required to realize an AI capability for hypersonic flight tracking, including getting the right constellations of space-based sensors in orbit, increased understanding of the flight characteristics of the missiles, and AI system training using sparse real-world data combined with simulated data. The full range of capabilities is not likely to be deployed soon, but an AI system could advance from a relatively simple deployment to a more sophisticated system as technologies are developed, tested, and put into service.

Regardless of whether the detection and warning system aims to detect only ballistic missiles or both ballistic and hypersonic missiles, a well-trained AI system, especially one processing data from a larger sensor array, could be better and faster than people at discriminating between incoming missiles and other objects or noise present in the environment or between a nuclear and conventional warhead.[86]

Finally, in addition to increasing the speed, robustness, and ability of early warning, AI may have other benefits in detection and sensor systems. For example, AI-enabled error pattern detection may provide information about a faulty warning system and whether or how it might be remedied.[87] Machine learning could be used to diagnose failures in sensor systems or to optimize data from a given number of sensors.

Prelaunch Detection

Using AI with advanced satellite-based imagery systems (likely different sensors than deployed for missile detection and warning described above) also creates the possibility to detect prelaunch activities for ground-mobile missiles that have been difficult to uncover. One unclassified news report suggests such a system is under development in the United States.[88] However, there is a possibility that countries being observed could develop "AI camouflage" to prevent AI from finding a hidden object more easily than hiding it from a human analyst.[89] Additionally, with continued advancement of sea-based sensors, it is conceivable that the ability to locate submarines could be developed, although there are still many technical obstacles to overcome before this is realized.[90] The potential to detect ground mobile missile launch preparation and/or location is a concern, especially to China and North Korea,[91] because they rely on ground mobile missiles for the survivability of their nuclear forces, although both are reportedly working on the capability for submarine-launched ballistic missiles (SLBMs). Russia also has a significant mobile ICBM force in addition to silo-based ICBMs and SLBMs. A sober analysis of mobile missile detection by Paul Bracken, an expert on the strategic application of technology in defense and business, suggests that AI along with other advancing technology will produce a new arms race.[92] The possibility of transparent oceans is of concern to military powers that rely on submarine-launched nuclear missiles for survivability, especially those that rely on narrow channels for departure or access.

Attack Characterization

The United States, and perhaps other states with early warning systems, already has sophisticated capabilities to characterize ballistic missile attack. In the United States, ground stations for IR satellites and ground radar installations analyze trajectories and targets and relay that information to NORAD. In some cases, other relevant information, such as data from other satellites or ISR information, is combined with ballistic missile tracking to form a complete operational picture and corroborate or question the incoming information. Experts expect that AI will continue to be explored for data fusion and to increase the speed and confidence of corroboration.[93]

In a case in which an IR satellite sensor detects launch from a site where launch preparation has been observed, attack characterization analysis could begin and/or a warning alert could be sent before confirmatory detection by other sensors is obtained. This would give leaders the maximum amount of decision time, and confirmatory signals and updated analysis could be relayed while decision making and communication were already underway. If the confirmatory signals did not validate the suspected trajectory or target, then the decision options could be adjusted in the face of the possibly that a false alarm had been triggered.

Benefit, Risk, and Strategic Stability Implications of AI in Early Warning Systems

There are two significant benefits offered by AI in early warning systems: the ability to quickly fuse and analyze data from early warning sensors and other military, commercial, or government sensors; and the potential to distinguish, in real time, nuclear weapons from conventional missiles, decoys, or unexpected phenomena or objects. If a greater number of sensors is deployed, AI could provide additional benefits, potentially including faster warning of ballistic or hypersonic missile approach and attack characterization, observation of prelaunch activity, and continuous assessment of system health.

A significant risk for AI application in early warning is that it generates erroneous information, as training data are relatively sparse and tests to obtain data are limited and expensive. It may be hard for a human-in-the-loop to notice an error in the AI-produced results or for humans to develop independent results in the needed time frames. To avoid decision making based on incorrect information, the AI systems deployed should be transparent and understandable as well as continuously tested, validated, and verified. Such transparent, understandable AI systems do not yet exist, and neither do test protocols for validation and verification.

Strategic stability has long been thought to improve when decision time—currently limited to about 15 minutes for an ICBM attack—is increased for countries that might consider launch under attack. Faster early warning systems are not the only or necessarily the best means to increase decision time,[94] and nonetheless, AI-enabled early warning is one method to increase decision time. Strategic stability also could improve when a more complete or accurate picture of attack is provided in a timely manner. Therefore, when AI is used with current early warning systems, it is likely stabilizing if the risks are managed.

An enhanced sensor system coupled with AI could be designed and deployed to detect, track, and predict the target of missiles more accurately and faster than today's IR satellite and ground-based radar systems.

However, when AI is used with advanced sensor networks, the ability to more rapidly and accurately provide early detection and warning and/or to observe prelaunch preparations each has complex strategic stability implications, and the combination of the two has additional stability implications.

An enhanced sensor system coupled with AI could be designed and deployed to detect, track, and predict the target of missiles more accurately and faster than today's IR satellite and ground-based radar systems. Although the improvement in speed or robustness of detection may be modest for ballistic missiles, it is a significant improvement for new hypersonic missiles that are detected by ground-based radar late in their trajectory. To examine the strategic stability implications of tracking hypersonic missiles, the reasons to have hypersonic missiles and the reason to fear hypersonic missiles must be considered. To date, the primary reason that Russia claims to be interested in nuclear-tipped hypersonic missiles is to evade U.S. missile defenses. Therefore, the ability of the U.S. to track hypersonic missiles would likely be seen by Russia as destabilizing because it would be viewed as an element of missile interception. However, the United States views the Russian hypersonic systems as a threat that delays early warning. Therefore, the U.S. ability to track hypersonic missiles would be seen as providing more early warning time and therefore would be stabilizing.

Alternatively, or in addition to an AI-enabled sensor system for detection and tracking, an AI-enabled advanced satellite-based imagery system could be designed to observe prelaunch preparation. Observing prelaunch preparation would give decision makers advance notice for response, and this attribute could be seen as stabilizing. However, the fear of prelaunch detection could cause countries with small arsenals to adopt a first use or launch-on-warning posture out of fear of losing their second-strike capability. These countries also could feel a need to expand their arsenals. Therefore, a prelaunch detection capability to observe mobile missiles would be seen as destabilizing.

If AI-enabled sensor systems were in place that both could improve the ability to track missiles and observe launch preparation or movement of mobile missiles, it could be seriously destabilizing. Each of the advances is leading to new capabilities and new fears about the ability to detect and intercept incoming missiles, to secure second-strike capabilities, and to increase the speed of detection and warning. Therefore, although there are some stabilizing attributes of AI-enabled new sensor systems, the combination of new nuclear delivery technologies with new sensors and AI technology could drive crisis and arms race instability.

A Race Toward Instability?

In a new and largely unchecked contest to try to protect themselves and outdo one another, the United States and Russia are poised to introduce a host of new technologies that could trigger a new nuclear arms race.

Here is the scenario: Hypersonic missiles are being developed by Russia to evade U.S. missile defense and provide them with a secure second strike. In response, the United States is considering the development of new sensor systems to better track missiles because hypersonic weapons complicate early warning, and the United States wants to avoid an attack that is undetected or detected with little time to act. The new U.S. sensor system may generate so much information that an AI capability will be needed to provide analysis to decision makers.

Consider the following implications and questions:

- The AI-enabled sensor system the United States may develop would be effective or perceived to be effective at both early warning and tracking missiles for interception. That might prompt Russia to seek additional ways to penetrate missile defense.
- Russia already is developing an underwater nuclear torpedo as an alternative to hypersonic glide vehicles. Will the United States respond by developing and deploying new underwater sensors? Will the underwater sensor systems advance to be able to detect submarines that carry nuclear weapons and threaten assured second strike from SLBMs?
- Will all states with nuclear weapons seek to deploy new sensors and AI? What are the new vulnerabilities, and how will they affect nuclear crisis stability?
- Will the missile tracking sensor/AI system prove so effective that an imaging version is developed for observing movement of land-based mobile missiles? Will that be destabilizing vis-a-vis Russia?
- Will China move to a launch-on-warning posture? Will China and North Korea further increase their stockpiles? What else might they do to assure second-strike capabilities? Are we already observing these changes?

Perhaps the most stabilizing element of AI with advanced sensors is the ability to distinguish an actual nuclear-weapon attack from a conventional attack or unexpected phenomenon or object. However, even this is nuanced as it also could improve missile defense by facilitating faster and more accurate differentiation between decoys and actual missiles.

Ultimately, AI application to early warning systems might improve strategic stability if used with today's space-based sensors and radars by combining collected data with other information to provide more comprehensive situational awareness and potentially more time and data for decision making. It also may allow for the ability to distinguish between nuclear and other types of missiles or objects. However, AI use in early warning systems that are enhanced by increasing the number and/or types of sensors will complicate the strategic stability calculus. Employing AI with advanced sensors for early warning should be done cautiously both to lower the risk of AI system failure and to ensure that strategic stability consequences are taken into account.

Decision Support

NC3 systems have three primary functions: situation assessment, course-of-action development and evaluation, and direction of force. These three functions often occur iteratively.[95] Early warning along with force management are the primary elements of situation assessment. Course-of-action development and evaluation includes nuclear planning and decision-making conferencing and requires the best decision-support tools possible. This section discusses how course-of-action development and evaluation could take advantage of AI by providing better decision-support tools, including tools to help with situational awareness (apart from early warning) and response options.

China has been specific about its interest in using AI technologies to support command decision making with decision-assistance systems.[96] The United States and Russia[97] have been explicit about developing decision-support tools, and both have employed situational awareness and response option tools to some extent in ongoing conventional battles.

Situational Awareness

An increasingly coveted use of AI in military applications involves providing situational awareness on the battlefield to support decision making.[98] AI is particularly good at situational awareness due to its ability to perform data mining (collect information), data fusion (combine information), front-line analysis (find trends, patterns, and anomalies), and predictive analysis (suggest new patterns).[99] Because ISR is increasingly being done with images and video collected from drones or other uninhabited vehicles, AI is especially useful to process the data quickly and highlight frames of interest for people to examine more closely. Project Maven, an AI experiment undertaken by the U.S. DoD in 2017, is one example of an AI application of this sort.[100] Additionally, situational awareness is becoming increasingly necessary due to the evolving nature of warfare in which, for example, a nuclear conflict could be preceded by, or conducted in parallel with, cyber info[101] and conventional conflicts. Other types of weapons of mass destruction, such as chemical or biological weapons, also could be deployed in certain cases. As a result, there is a growing interest among military leaders in using the best situational awareness tools available for hybrid conflict scenarios.

In 2020, Gen. John Hyten, who serves as vice chairman of the Joint Chiefs of Staff, said that the updated U.S. NC3 and the new conventional forces Joint All Domain Command and Control (JADC2) system will be interfaced: "NC3 will also operate in things that are separate from JADC2 because of the unique nature of the nuclear business, but it will operate in significant elements of JADC2. Therefore, NC3 has to inform JADC2 and JADC2 has to inform NC3."[102] It is widely believed that JADC2 will incorporate AI to provide situational awareness in time frames that are relevant.[103] Whether or not AI is used directly in U.S. NC3, there likely will be an interface to a system that is employing AI for situational awareness. Also, as mentioned in the preceding section on early warning, AI already may be in development or early use for fusing data from early warning and other government and commercial satellites.

It is not clear whether AI is needed to develop response options beyond situational awareness. A RAND study concluded that AI is less likely to be implemented directly into NC3, but computer programs, simulations, and data analysis can still quickly provide options for human decision makers to consider.

In Russia, the Moscow-based National Defense Control Center (NDCC) was inaugurated in 2014 as a centralized command post for assessing global threats and initiating whatever military action is deemed necessary, whether nuclear or non-nuclear. The NDCC is designed to collect information on adversary moves from multiple sources and provide senior officers with guidance on possible responses.[104] The NDCC is touted to host a supercomputer made only with Russian computer components with three times more computing power and 15 times more storage capacity than the Pentagon.[105] More recently, Russia has announced that AI will be introduced for decision making during an NDCC upgrade.[106]

China has revealed systems aimed to fulfill a range of operations, including joint theater operations and global surveillance and strike. These systems provide multisource fusion and comprehensive integration of information and are intended to eventually provide all-domain operations. Due to the scarcity of open-source information on China's military data collection and analysis system, it is difficult to assess its capability or maturity.[107]

Response Options

It is not clear whether AI is needed to develop response options beyond situational awareness. A RAND study concluded that AI is less likely to be implemented directly into NC3, but computer programs, simulations, and data analysis can still quickly provide options for human decision makers to consider.[108] As mentioned earlier in this paper, the 2018 Nuclear Posture Review said that advanced technologies would be explored to support decision makers, and such advanced technology could be as obvious as display techniques to visualize possible courses of action or as esoteric as using AI technologies to predict the risks involved in possible response scenarios.

Risk investigation requires determining the impact of military action, whether nuclear or non-nuclear, and possible adversary responses and their impact through repeated cycles. For example, in a nuclear response scenario, the impact could be determined through calculation of the consequences of nuclear use, including blast, radiation, and other effects, with real-time parameters, such as weather and population densities along with target characteristics. Such calculations are very sophisticated but would not necessarily need AI. However, other elements of the response scenario might benefit from AI. For example, a sophisticated

and well-trained AI system could play against itself to predict the next several steps that could occur in a conflict. If AI training like this is possible and produces reliable results, the AI system could help decision makers by performing risk calculations.[109]

Benefits, Risks, and Strategic Stability Implications of AI in Decision Support

AI-aided situational awareness could provide a more complete and timely picture of the operating environment. This knowledge could be stabilizing because it could alleviate pressure to use nuclear weapons if the system determined, for example, that a massive attack was not underway but panic was occurring through social media escalation. Risk-informed, AI-aided response-option generation would go further by providing not only situational awareness but evaluation of how actions might affect conflict escalation and therefore provide clearer escalation-ladder options to avoid a nuclear exchange.

However, there is a risk that incorrect analysis could be elevated from conventional force situational awareness tools to a modern NC3 system. Approaches that quickly and thoroughly validate information exchanged between systems is essential. It also is possible that algorithms could mischaracterize whether actions are linked or independent, so learning and understanding human-machine partnerships and trust relationships would be essential. Additionally, people would require new training with these systems to avoid automation bias or destabilizing actions while benefiting from their analytical power. If carefully implemented, AI-enabled situational awareness could solve one of the fastest-growing challenges today: processing all available information in a short time to make good decisions.

Automated Retaliatory Launch

The final element of NC3 is direction of the force that raises the issue of whether an automated retaliatory launch option is desired. The specter of automated nuclear weapon retaliatory launch has been viewed simultaneously as risky and potentially stabilizing—risky because weapons could be launched by machine error and potentially stabilizing because of the lower likelihood that a nuclear attack would be initiated if retaliation was known to be automated, swift, and certain. That said, U.S. decision makers have determined that the devastating consequences of accidental nuclear war outweigh reliance on retaliatory launch without people in the loop.

However, there have been three primary considerations that keep automated retaliatory launch, with or without a human-in-the-loop, as an option that some states with nuclear weapons might see as beneficial:

1. **Concern that a leader with sole authority to authorize nuclear launch could be killed or lose contact before a retaliatory strike is ordered.**

   The fear of losing communication with the leader(s) responsible for authorizing a nuclear counterstrike was the motivation behind the only known automated nuclear retaliatory launch system, Perimeter, which was deployed in 1986 in the Soviet Union. Relatively recent publications by Russian authors have somewhat clarified the Perimeter system.[110] It was designed to be used if a network of sensors detected nuclear detonations in the Soviet Union and communication with Soviet leadership was lost. Unless deactivated in a specified period, Perimeter would transfer nuclear launch authority to duty officers in an underground bunker who could launch communication rockets that beam launch codes to hardened missile silos.[111] Current descriptions report that Perimeter is now

the Russian back-up NC3 system to the primary NC3 system "Kazbek," the Combat Management Automated System of the Nuclear Strategic Forces. Perimeter's capabilities and characteristics are unknown, but some unconfirmed media reports claim it uses some level of AI.[112]

An alternative approach to automated retaliatory launch, called "decide under attack," has been proposed by retired Admiral James Winnefeld.[113] In this proposal, the U.S. president would decide while missiles were thought to be in flight what he or she wanted done if nuclear explosions began and he or she were killed. The response would not be conducted until multiple nuclear explosions were confirmed and a number of hours had elapsed.[114]

2. **Concern that a nation's second-strike capabilities are at risk from a first strike and must be launched quickly before being struck.[115]**

   If a state perceives its nuclear arsenal to be at risk from an adversarial first strike, a leader may be required to make a "use or lose" decision, especially if their nuclear force is small and/or not hardened. Second-strike vulnerability could increase as sensors and AI systems advance (as discussed in the previous section on prelaunch detection), raising concerns that mobile ICBMs and perhaps SLBMs could become more vulnerable to attack.[116] States that do not have sophisticated NC3 or states that may be concerned about their second-strike capability may choose to develop automated retaliatory launch to act quickly as opposed to relying on vulnerable communication channels to ensure fast response.

3. **Belief that the speed of war requires automated response.[117]**

   Military educators Adam Lowther and Curtis McGiffin argue that recent technological trends add a new justification for considering automated retaliatory launch capabilities. They recommend the United States "develop a system based on artificial intelligence, with predetermined response decisions, that detects, decides, and directs strategic forces with such speed that the attack-time compression challenge does not place the United States in an impossible position."[118] This argument highlights the inherent tension between human-in-the-loop policies and AI's main value added—its automatic, faster-than-human response to threats.[119]

These considerations suggest automated retaliatory launch will not be completely dismissed despite its recognizable dangers. In the first two scenarios described above, an AI system would help ensure launch. In the third scenario, Lowther and McGiffin suggest targets would be predetermined; however, an AI system can be envisioned that would both help ensure launch and dynamically select targets. These capabilities present serious risks as well as potential benefits.

Assured Launch

For automated nuclear weapon launch against a preplanned target(s), an AI system could be used to control the sequence and timing of events, with or without human involvement. AI also potentially could provide a redundant system in case of unexpected actions from a human-controlled system or offer escape options if something doesn't go as planned.

Target Selection

It is imaginable over the longer term that an AI system could be used to optimize target selection just prior to launch. Such a system would rely on real-time early warning and/or other sensor information to select either the attack plan most likely to be successful (for example, not likely to be defeated by missile defenses) or the attack plan that would achieve the most important military objectives. This type of target analysis also is useful in a decision-support system that provides options for people to consider. An AI system does not require decision makers to contemplate the realities, it simply works to optimize the attack as quickly as possible.

Benefits, Risks, and Strategic Stability Implications of AI in Automated Retaliatory Launch

Automated retaliatory launch has the highest risk of all the AI applications to NC3 because failure could result in a nuclear attack. The risk seems extraordinarily high for a system without a human-in-the loop, even if such systems have the full benefit of machine speed and autonomy. It is difficult to justify why a nuclear war would be better fought at machine speed, especially because nuclear weapons are meant primarily as a deterrent.

Even an AI-enabled automated launch system with a human-in-the-loop has significant risks. Critics argue that the proposal for a U.S. "Dead Hand" would increase the risk of nuclear use due to automation bias and unknown problems that may be associated with simulated data.[120] Vulnerabilities within Russia's Perimeter system could lead to high-consequence mistakes. For instance, risk researcher and RAND contributor Anthony Barrett suggests that Perimeter could misconstrue and misattribute data caused by a meteorite strike to detonations from a U.S. nuclear attack or that the system could launch nuclear weapons at the United States if communication lines were severed during a terrorist attack.[121]

Automated retaliatory launch has the highest risk of all the AI applications to NC3 because failure could result in a nuclear attack.

Summary of AI in NC3 Systems

There are many possible uses of AI in modernized NC3 systems. It seems inevitable that some of these applications will be considered by one or more states with nuclear weapons and will likely be implemented when the benefits are perceived to outweigh the risks. Hopefully, before AI systems are implemented, research will have been done to improve transparency and explainability; lower the risk of failure; establish robust testing, validation, and verification procedures; and engineer fail-safe approaches. Ideally, the technical risks will also be lowered by agreement on international norms and standards before significant AI implementation.

The AI applications that seem most likely to be implemented by well-resourced states with nuclear weapons are cybersecurity and decision support (including within early warning systems and between nuclear and conventional command and control) because they could improve complex tasks and enhance strategic stability with manageable risks. AI in early warning systems with enhanced sensor networks and automated retaliatory launch have multifaceted stability considerations, and automated retaliatory launch could have grave consequences if an error occurs. International norms of behavior or policy declarations about the use of AI in nuclear-weapon systems would be beneficial to increase transparency and manage perceptions even in the applications with the potential to improve stability and with relatively low risks, such as cybersecurity.

Autonomous Nuclear Weapon Systems

Lethal autonomous weapons are weapon systems that can identify, select, and engage a target without meaningful human control.[122] Such weapons are software controlled and do not maintain contact with human operators allowing them to act on their own and potentially maintain longer operational lifetimes without risking communication jamming or interference.

By this definition, there are no lethal autonomous nuclear weapons or associated delivery vehicles today. For instance, even though ballistic missiles do not communicate with humans once launched and, in some cases, have a modest level of autonomy to maintain their ballistic flight trajectories, they are not fully autonomous weapon systems because they have preselected and preprogrammed targets and a predetermined course. Similarly, aircraft and submarines that carry nuclear weapons are human-operated nuclear-weapon delivery systems from which nuclear weapons are launched with a preselected target and predetermined trajectory.

To be considered fully autonomous, a nuclear-weapon system would need to have capabilities that don't exist in today's nuclear-weapon systems or in many other military or commercial systems. Full autonomy would include the ability to change trajectory, detect and avoid obstacles or interceptors, automatically identify a target, loiter, or return "home" without human engagement.

Although technical progress in autonomous vehicles has been impressive in both the commercial and military sectors, significant hardware and software challenges remain. These include the need to develop energy-efficient computers that store large amounts of data and are very fast and latency-free communication pathways between sensors and computational systems to enable real-time decision making. Despite advances, today's algorithms still face challenges reliably identifying objects, responding in real-time, planning and controlling routes in the absence of GPS, and defending against cyberattacks.[123] These same challenges exist to different degrees for missiles, drones, robots, and self-driving vehicles. As a result, AI military research is underway to improve important elements such as navigation, guidance, and targeting. For conventional missile systems, research aims to allow "terrain-hugging" without collisions.[124] Because of the technical challenges, systems in use today in both the commercial and military sectors tend to be semi-autonomous (remotely controlled or human-assisted) rather than fully autonomous.

Semi-Autonomous Nuclear Weapon and Military Activities in the United States, Russia, and China

The following section describes activities in the United States, Russia, and China on semi-autonomous military and nuclear-weapon systems and analyzes the potential benefits, risks, and stability implications if semi- or fully-autonomous nuclear-weapon systems were deployed.

United States

All statements to date are clear that the United States does not plan to introduce autonomy into nuclear-weapon delivery vehicles or nuclear weapons. There is a plan to have optional autonomy for the future nuclear-capable B-21 heavy bomber,[125] but only non-nuclear missions will be autonomous; all nuclear missions will be under human control.[126] The B-21 Raider is being designed to have sufficient on-board reconnaissance, targeting, and self-defense features to accomplish missions in the most hostile environments.

The non-nuclear military applications of autonomy in the United States provide a sense of technology maturity and direction. The United States has deployed drones for ISR and precision strike missions with human-in-the-loop control. Currently, the United States is developing autonomous or optionally human-controlled aerial vehicles including electronic warfare aircraft, supply helicopters, wingman aircraft, as well as the long-range bomber. The United States is also developing sea-based autonomous systems including mid- and large-scale surface ships, and underwater vehicles, including extra-large submarines.[127] Additionally, semi-autonomous ground vehicles have been deployed to follow soldiers or vehicles and accomplish independent tasks. In the future, these ground vehicles are intended to be more—and perhaps fully—autonomous.[128] The United States is conducting research for applying AI to hypersonic missiles for more sophisticated autonomy.[129] Multiple U.S. programs aim to develop swarming drones—both air and underwater. The swarming initiatives may force the creation of fully autonomous military systems because they cannot effectively swarm and be controlled by human operators.[130] The autonomous systems being developed in the United States are aimed at far less consequential applications than nuclear-weapon systems.

Due to the high consequences of failure of a nuclear-powered cruise missile carrying a nuclear weapon, the autonomous system must be particularly robust and well tested, and it would have to be able to fail in a safe manner.

Russia

President Putin announced in March 2018 that Russia was developing two autonomous nuclear-weapon delivery vehicles: a nuclear-powered underwater vehicle called Poseidon (sometimes called Status-6 or Kanyon) and a nuclear-powered cruise missile referred to as "Burevestnik."[131] The degree of autonomy intended for these systems remains unclear and undisclosed. The Russian nuclear hypersonic boost-glide vehicle, Avangard, could have some limited autonomy but is not believed to have full or sophisticated autonomy at this time.

It is unclear whether the Poseidon autonomous underwater vehicle would have a preprogrammed underwater route or would be capable of detecting and maneuvering around natural or man-made underwater obstacles in its long-distance course without human control. If AI is used in its navigation, the issue of power-intensive computing may be alleviated because it is powered by a compact nuclear reactor. However, significant sensors, training data, and algorithm development would be needed and would make the weapon significantly more flexible. Regarding Poseidon's ability to locate and track targets, its large nuclear yield suggests that it is not meant to be a precise weapon whether used on U.S. infrastructure, coastal cities, or carrier groups. Therefore, it is likely launched with a predetermined detonation location. Although not verified, it appears the Poseidon is probably semi-autonomous.

The Burevestnik cruise missile is generally believed to be in an earlier phase of development than Poseidon due to the additional difficulties of developing a nuclear power system and an associated engine that is lightweight, compact, and safe enough to fly. Burevestnik was announced as having the ability to stay aloft for days and potentially loiter or change course. Sensors, data, and algorithms will be needed for safe autonomous flight. Due to the high consequences of failure of a nuclear-powered cruise missile carrying a nuclear weapon, the autonomous system must be particularly robust and well tested, and it would have to be able to fail in a safe manner. There is no open-source detail about the autonomous technology for Burevestnik.

Russia's non-nuclear military drones and autonomous vehicles are generally perceived to lag U.S. and Chinese development. However, as displayed at the Army-2020 Forum in Moscow, Russia has mounted a concerted effort to catch up on attack and ISR drone capabilities.[132] Furthermore, Russia has focused autonomy investments on systems that can help soldiers counter the physical, cognitive, and operational challenges of urban warfare.[133] Of note is the use of unmanned ground vehicles, ISR drones, and situational awareness tools. Russia also has reportedly built a combat module for unattended ground vehicles that is capable of autonomous target identification—and potentially, target engagement—and plans to develop a suite of AI-enabled autonomous systems.[134] In Syria, Russia has battle-tested unmanned aerial vehicles for electronic warfare and new night-sensor systems. Finally, Russia remains interested in exoskeleton warfighting robots. Russia appears to be focusing autonomy investments on urban warfare and nuclear-weapon delivery vehicles.

China's interest in military AI and autonomy is wide-ranging and documented in a number of published plans. However, regarding autonomy applied to nuclear weapons, China's intent is unstated.

China

China's interest in military AI and autonomy is wide-ranging and documented in a number of published plans.[135] However, regarding autonomy applied to nuclear weapons, China's intent is unstated. There is wide speculation that China's DF-ZF hypersonic glide vehicle could be dual-capable and at least partially autonomous, although neither has been confirmed by Chinese officials.[136] The DF-ZF has been tested and displayed with the DF-17 medium-range ICBM and there is speculation that the DF-ZF could be carried by other medium- or long-range ICBMs.

Because of the lack of open-source information regarding China's nuclear program, it is necessary to assess potential autonomy in nuclear-weapon systems by exploring other military development. Activities in China that might be dual use include AI use to enable (1) improved autonomy in control and targeting of cruise missiles, (2) more precise and autonomous control of hypersonic glide vehicles, and (3) autonomous flight control of long-range unmanned aerial systems.[137] In addition to the DF-ZF hypersonic glide vehicle, there is some evidence to support the use of AI for autonomy in the GJ-11 Sharp Sword stealth combat air vehicle, the CJ-20 air-launched cruise missile, the YJ-100 subsonic anti-ship missile, the HSU-001 extra-large underwater vehicle, and the CJ-10 land attack cruise missile.[138]

China has engaged in significant efforts in asymmetric autonomous operations in the space and cyber domains, as well as autonomous vehicles and swarming. Research on various types of air, sea, undersea, and land autonomous vehicles is ongoing, with a specific interest in swarming capabilities. An AI-enabled swarm of more than 1,000 vehicles was demonstrated in a 2017 Chinese airshow, and a media report after the demonstration showed a computer simulation of a large swarm finding and destroying a missile launcher.[139] In 2021, a swarm of more than 3,000 vehicles was demonstrated.[140]

Benefits, Risks, and Stability Implications of Autonomous Nuclear Weapon Systems

At this time, the greatest operational benefit provided by autonomy for nuclear-weapon systems appears to be maneuverability to provide longer unmanned routes while avoiding missile defenses or obstacles. Over time, more interest could develop in autonomously identifying and pursuing mobile targets such as carrier groups.

However, autonomous nuclear weapons or delivery systems come with significant risk. Autonomous systems today are being developed in parallel with or before establishing test, evaluation, validation, and verification protocols; risk evaluation approaches; or fail-safe techniques. In addition, autonomous systems could be at higher risk of manipulation through adversarial cyber hacking or spoofing. Cyberattacks could target navigation sensors or communication paths.[141] Spoofing by sending false GPS coordinates is known to be able to alter drone flight path,[142] and GPS- or radar-jamming technologies are cheap and technologically simple.[143] States with nuclear weapons will have to decide on the acceptable risk of introducing AI as the technology matures. States willing to assume more risk may introduce AI in their systems sooner.

A potentially stabilizing element of autonomous nuclear-weapon delivery systems may be increased confidence in secure second strike due to the ability to go undetected by maneuvering during delivery, for example, using ground-hugging techniques or deep undersea routes. Such maneuverability also increases the cost of an adversary's missile defense system. Additionally, nuclear-weapon systems that can more precisely and dynamically locate targets could reduce the requirements for number and yield of nuclear weapons needed to achieve the intended deterrent or military effects.

The unpredictability of autonomous nuclear-weapon systems tends to offset any potential stabilizing benefits, however. With maneuverability, longer range, and features such as loitering, adversaries will not be able to predict attack characteristics and may be more likely to assume the worst and potentially escalate. Combined with risks associated with cyberattacks or other failures, fully autonomous nuclear-weapon systems are particularly worrisome.

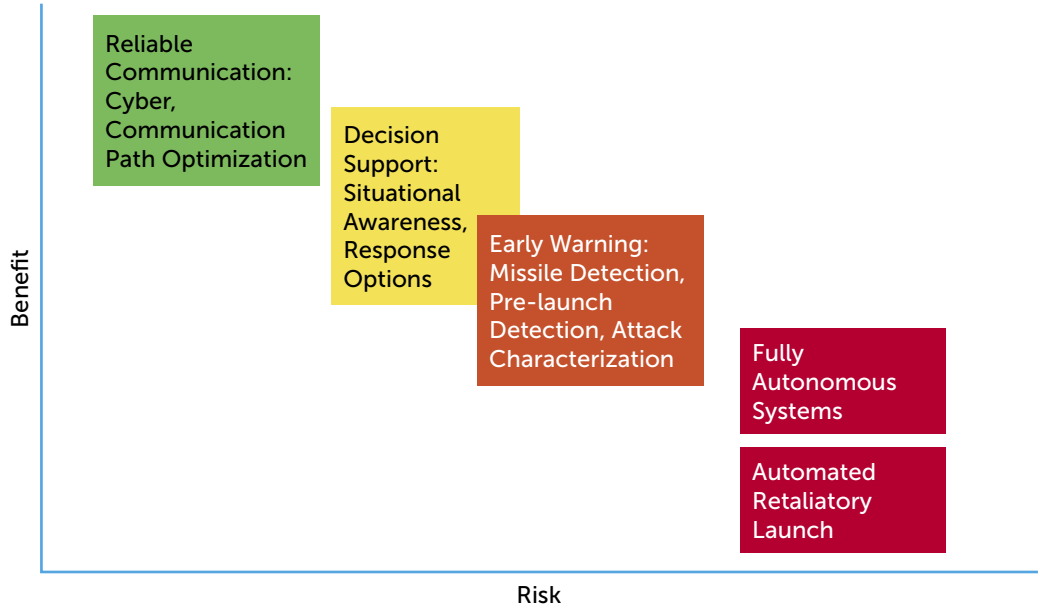
Although it is impossible to predict whether, when, or how AI-based autonomy will be introduced into nuclear-weapon delivery systems or nuclear weapons, it is likely that improved maneuverability and autonomous target recognition will be perceived as beneficial and at least partially incorporated over time. Fully autonomous nuclear-weapon delivery systems or nuclear weapons will take longer to develop and are much riskier to deploy. Fully autonomous systems may become more attractive if deployed air- and space-defenses improve substantially and/or as stockpiles become smaller.

Given the implications for stability and increased nuclear risks in fully autonomous systems, a ban on fully autonomous systems should be considered. In a ban agreement, provisions could be provided to allow certain semi-autonomy as the technology is better understood and proven. Nuclear policy experts Jessica Cox and Heather Williams[144] have suggested banning or limiting fully autonomous, nuclear-armed weapons systems that cannot be recalled or redirected in a crisis. This proposal based on recallability may deserve consideration as technology matures and is better understood, but until then it seems prudent and likely more verifiable to agree to a ban on fully autonomous systems.

Comparative Analysis of AI Applications to Nuclear Weapons and Their Operational Systems

Figure 4 provides a comparative visual representation of the risks, benefits, and strategic stability implications of the various AI applications to nuclear weapons and their operational systems. As seen in the figure, in NC3 systems, the most beneficial and stabilizing application of AI with the least risk is to employ it to ensure reliable communication; the highest-risk and most destabilizing application is related to automated retaliatory launch. Fully autonomous nuclear-weapon systems also are seen as highly risky and destabilizing, although some operational benefits could be gained. AI-supported decision-support systems and early warning systems have real benefits but their risks and stability implications should be closely considered if they are developed for use in nuclear-weapon systems. Any application of AI needs more research, test, and evaluation to lower the risk of failure before use in nuclear-weapon systems, and verification, validation, and fail-safe approaches need to be developed.

Figure 4: Relative Benefit, Risk, and Impact on Strategic Stability of AI Applications to Nuclear Weapons and Their Operational Systems



Note: The impact on strategic stability is indicated by the color of the application box, with green being the most likely to have a stabilizing effect, yellow and orange indicate careful consideration for impact on stability, and red is most likely to have a destabilizing effect.

AI Applications in Other Areas Related to the Nuclear Enterprise

Beyond the AI applications that directly interface with nuclear weapons and their operational systems, there are potential applications in other areas that support or relate to nuclear weapons and the nuclear enterprise. Of particular note are AI applications to war gaming and planning, physical security systems for nuclear weapons or special nuclear materials, missile defense, and monitoring nuclear proliferation and arms control agreements. These four areas are not covered in detail as they have less impact on nuclear strategic stability.

A significant benefit for AI could be realized in monitoring nuclear proliferation and arms control agreements. Other areas with clear benefits and manageable risk are war-gaming and planning and physical security systems. AI application to missile defense has benefits, but it could escalate an offense-defense arms race.

AI for War-Gaming and Planning

War-gaming supported by advances in computer graphics and AI has been prioritized by both U.S. and Chinese military leadership.[145] Although AI can enhance the immersive experience in war-gaming, tailor games to individuals, and record actions for learning objectives, its real power for war-gaming and planning is to test new AI tools in a simulated operational environment, train with new AI tools, and explore human-machine interactions in data-driven decision making and operations. Introducing AI-on-AI games also can assist the development of new concepts, risk calculations, and strategies for planning purposes. Ideally, AI incorporation into war-gaming and planning will reinforce the high consequence of nuclear use.

AI for Physical Security of Nuclear Weapons and Materials

Physical security systems used to protect nuclear weapons and special nuclear materials from theft or sabotage need to continuously improve and adjust to changing security threats, such as cyber incursions and overhead drone attacks.[146] Innovative technology also is needed to improve the efficiency and responsiveness of guard forces.

AI could provide beneficial new capabilities to help monitor the perimeter and interior camera, video, and sensor systems commonly used for physical security to trigger indications of potential concern to the security officers. AI could also continuously assess the integrity of the security sensor networks to identify and respond quickly to restore the network in the event of ordinary failures or a deliberate attack. AI-enabled drones and robots could either augment or replace human patrols of a perimeter. Semi-autonomous or autonomous response systems, from guns to robots to vehicles, could supplement or replace human response forces in the event of an intrusion. In these various ways, AI-enabled systems could reduce the number of security forces needed, improve the effectiveness of security forces in place, and/or detect and respond to threats that are difficult to counter today.

If well implemented, AI has a large potential benefit to improve physical security systems. Unfortunately, AI also introduces new attack vectors, especially for cyber hacking, so its implementation needs to be continuously tested, validated, and verified.

AI for Proliferation Monitoring and Arms Control Agreements

The ability of AI systems to digest and evaluate a large amount of data from a variety of sources is a potentially game-changing technology breakthrough for monitoring nuclear proliferation and arms control agreements.

For example, data from commercial space-based sensors with optical (visible or infrared) image, video, or change-detection capabilities have increased significantly. The spectral, spatial, and temporal resolution also has improved. If these data are continuously analyzed by an AI system, nuclear proliferation activities are likely to be detected earlier.[147] A RAND working group considered that sensor networks and AI could enhance transparency and trust for arms control treaty monitoring, and "a future AI system could essentially be the arms control regime, monitoring compliance and adjudicating violations without human input."[148]

In addition to the vastly increased information from spaced-based sensors, data from various sources can be evaluated and combined to create a more complete picture of potential nuclear proliferant activity. In 2019–2020, NTI partnered with Center for Advanced Defense Studies (C4ADS) to examine trade data to potentially indicate covert nuclear proliferation activities.[149] The study, released in early 2021, found that AI could be used with public domain trade data to expose otherwise unknown companies potentially involved in proliferation transactions.

In addition to open-source data, governments collect data used for national assessments. The national data can be merged or compared with open-source data for corroboration. A significant benefit of open-source information is that it can be used to make a public case against a proliferator, whereas government-collected data cannot always be released out of concern it could disclose sources or methods. Research will be needed to create, test, validate, and verify systems and to ensure proper interpretation of results. However, the potential benefits are compelling and likely to increase with time.

AI for Missile Defense

Missile defense systems continue to be deployed around the world and are now a significant driver for new offensive nuclear capabilities.[150] Because missile defenses are expensive, AI might be seen as a way to improve effectiveness.

There are at least three ways that AI can help optimize missile defense objectives:

- Use sensor data to quickly predict the trajectories and targets of incoming missiles. Data from the new sensors being considered for early warning systems also could be used for missile defense applications.
- Optimize missile defense assets in real time, which would require the aid of accurate tracking.[151] This application could help lower the number and cost of missile interceptors.
- Aid missile defense by improving automatic target recognition on interceptors using machine learning techniques.

Although there are benefits to missile defense systems using AI, even the perception that research is underway on AI-enabled tracking and targeting capabilities, optimization schemes, or target recognition could be destabilizing. RAND researchers Edward Geist and Andrew Lohn note that despite technical limitations of AI-enabled tracking, the system needs only to be perceived as capable to be destabilizing.[152]

Recommendations

In the absence of policy decisions to the contrary, it appears likely that as technology advances, AI will be incorporated into nuclear and related systems because it provides benefits that will be seen as outweighing the associated risks. In some cases, AI might simply provide a reliable and automated approach (cybersecurity or situational awareness), and in other cases, it might be required for entirely new capabilities (decision support or autonomous delivery). For each state with nuclear weapons, the specific application and timing of AI incorporation will depend on production or modernization schedules, the perceived benefits and risks, the technical capabilities and level of investment required, and the importance of AI safety compared to the perceived benefits of deployment.[153] The analysis in this paper leads to the following three recommendations.

1 Research Is Needed to Lower Technical Risk and Develop Fail-Safe Protocols

AI research around the world is significant and includes important topics discussed in this paper, such as explaining ML results, understanding the cyber offense-defense relationship, preventing algorithm manipulation, optimizing human-machine partnership, and learning with simulated data. Advances in understanding the implications related to these topics are essential for considering AI incorporation into nuclear-weapon systems. In addition, significantly more research and development aimed at operationalizing AI systems with low technical risk in high-consequence applications is needed, along with work to define and engineer fail-safe systems.[154] The U.S. national security enterprise should prioritize research in this area.

These research and development activities will need to include the development of entirely new test, evaluation, validation, and verification protocols;[155] development of a framework for evaluating risk; technical innovation to lower risks; and methodologies to ensure fail-safe operation.[156] Paul Scharre, an expert in the military use of autonomy, warns that "a rush to field AI systems before they are fully tested could result in a 'race to the bottom' on safety, with militaries fielding accident-prone AI systems."[157]

Research ideally should engage the best minds in the U.S. commercial and military sectors as well as academia and Federally Funded Research and Development Centers. Because states with nuclear weapons are most likely to act responsibly if they have full understanding, open publication of the research should be encouraged when it does not compromise national security. Similarly, the United States should encourage international research collaboration in areas of mutual interest to produce high-quality results and develop relationships between people, institutions, and countries. It is not expected that the collaborative research would involve specific military applications. The United States should encourage other states with nuclear weapons to pursue similar research.

2 States with Nuclear Weapons Should Adopt Policy Positions Regarding AI

Because of the rapid development and disruptive potential of AI, states with nuclear weapons should consider declaratory policy statements that clarify their intent on human/machine control of nuclear weapons. These policy positions and declaratory statements would be responsive to heightened concerns and interest in AI across every sector, including nuclear-weapon systems. They could be quite simple, such as the following:

- Nuclear weapon use will always be authorized by humans.[158]
- Fully autonomous nuclear-weapon systems will not be developed.
- AI-decision-support systems within NC3 will always generate response options that include non-nuclear response as well as nuclear response.

3 International Dialogue on AI Use in Nuclear Weapons Should Be a Priority

AI is a topic that should be included in international dialogue on strategic stability at the expert level and the military level. Nuclear policy experts Jessica Cox and Heather Williams have suggested the Permanent Five members of the United Nations Security Council (China, France, Russia, the United Kingdom, and the United States) should add the benefits and risks of emerging technology such as AI as a standing topic on their annual agenda.[159]

The dialogue should include technical risks, strategic stability implications, and crisis stability implications. Perhaps cooperative efforts could begin in applications such as physical security systems and nuclear proliferation monitoring. Another potential area for cooperation could involve establishing international technical standards for test and evaluation protocols. Beyond dialogue and cooperation, agreements on the limits of AI in nuclear systems are worth pursuing.

As a rapidly emerging and enormously powerful technology, it is inevitable that AI will be considered for use in nuclear-weapon systems in addition to other military applications.

There are no easy answers for how AI should intersect with nuclear weapons. However, this is the right time to expand research efforts, consider policy approaches, and engage in constructive international dialogue regarding the risks and benefits of AI with respect to nuclear-weapon systems and strategic stability.

About the Authors

Jill Hruby was the inaugural Sam Nunn Distinguished Fellow at NTI from November 2018 to November 2019 and continued working part time with NTI as a Distinguished Fellow from November 2019 to July 2021. She was confirmed in July 2021 to be the under secretary for nuclear security in the Department of Energy. From July 2015 to May 2017, Hruby was the director of Sandia National Laboratories, the culmination of a 34-year career at Sandia working in nuclear weapons, nuclear non-proliferation, biological defense, homeland security, and science and technology fields.

This paper was initiated while Hruby was the Sam Nunn Distinguished Fellow and completed during her period as a Distinguished Fellow. In addition to her work with NTI, from 2017 until her confirmation as under secretary, Hruby contributed to the National Academies of Sciences, Engineering, and Medicine on the Committee for International Security and Arms Control, including chairing the study "Nuclear Proliferation and Arms Control Monitoring, Detection, and Verification: A National Security Priority (Interim Report)." In addition, she served on multiple boards and advisory committees and is a spokesperson for women in engineering.

Hruby holds a bachelor's degree from Purdue University and a master's degree from the University of California, Berkeley, both in mechanical engineering.

M. Nina Miller was an intern at NTI during the summer and fall of 2019. She is currently a PhD student in security studies and international relations at the Massachusetts Institute of Technology. Her research focuses on the intersection of security, technology, and political psychology, including strategic decision making and autonomous weapons. Miller holds a bachelor of arts (International Honours) in international relations from the Joint Degree Programme between the College of William & Mary and the University of St Andrews, and she graduated summa cum laude with first-class honors.

Acknowledgments

The authors wish to thank Ian Singer, a 2019 NTI summer intern from Stanford University, for his input during the early phases of conceiving and writing this paper. They would also like to thank Ernie Moniz, NTI's chief executive officer and co-chair, for his encouragement in evaluating both the benefits and the perils of technology for nuclear-weapon systems. A special thanks to Joan Rohlfing, NTI's president, for encouraging and supporting work on artificial intelligence and its intersection with nuclear weapons from the early days of considering this project. Lynn Rusten and Page Stoutland provided extensive and thoughtful comments on the drafts that helped shape the final paper and recommendations. The authors are also grateful to Mimi Hall of NTI's communications team.

Endnotes

[1] See Shana Lynch, "Andrew Ng: Why AI is the new electricity," *Stanford News,* news.stanford.edu/thedish/2017/03/14/andrew-ng-why-ai-is-the-new-electricity, and Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 1, no. 3, May 2018, p. 41, tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power.

[2] See Ryan Browne, "A.I. could lead to a nuclear war by 2040, think tank warns," CNBC, April 25, 2018, www.cnbc.com/2018/04/25/ai-could-lead-to-a-nuclear-war-by-2040-rand-corporation-warns.html. Browne quoted Elon Musk regarding an AI "immortal dictator from which we can never escape." See also Will Knight, "Why artificial intelligence might trigger a nuclear war," Ethical Tech, *MIT Technology Review,* April 24, 2018, www.technologyreview.com/f/610996/why-artificial-intelligence-might-trigger-a-nuclear-war. Knight cautioned that "well before Terminator robots rise and attack us, AI could help us destroy ourselves with nuclear weapons."

[3] See U.S. Department of Defense Release "DOD Adopts Ethical Principles for Artificial Intelligence," February 24, 2020, www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence, and "Beijing AI Principles," May 28, 2019, www.baai.ac.cn/news/beijing-ai-principles-en.html.

[4] Autonomous weapon systems are one application that requires AI. AI and autonomous systems are not synonymous. Autonomous weapon systems require technologies in addition to AI to be successful. For more information on the UN lethal autonomous weapons activity, see Alexandra Brzozowski, "No progress in UN talks on regulating lethal autonomous weapons," EURACTIV, November 22, 2019, www.euractiv.com/section/global-europe/news/no-progress-in-un-talks-on-regulating-lethal-autonomous-weapons.

[5] Kelley Sayler, "Artificial Intelligence and National Security," Congressional Research Service, updated August 26, 2020, R45178, fas.org/sgp/crs/natsec/R45178.pdf.

[6] Bernard Marr, "The 4 Ds of Robotization: Dull, Dirty, Dangerous and Dear," *Forbes*, October 16, 2017, www.forbes.com/sites/bernardmarr/2017/10/16/the-4-ds-of-robotization-dull-dirty-dangerous-and-dear/?sh=b40ba8c3e0df.

[7] James Vincent, "Putin says the nation that leads in AI 'will be the ruler of the world,'" The Verge, September 4, 2017, www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world.

[8] Elsa Kania, "Chinese Military Innovation in Artificial Intelligence," testimony before the U.S.-China Economic and Security Review Commission Hearing on Trade, Technology, and Military-Civil Fusion, June 7, 2019, Center for a New American Security.

[9] Franz-Stefan Gady, "Elsa B. Kania on Artificial Intelligence and Great Power Competition," *The Diplomat,* December 31, 2019, thediplomat.com/2020/01/elsa-b-kania-on-artificial-intelligence-and-great-power-competition.

[10] Stanley Center for Peace and Security, United Nations Office for Disarmament Affairs, and Stimson Center, "The Militarization of Artificial Intelligence," workshop held August 2019 at United Nations New York, workshop summary published June 2020, www.un.org/disarmament/the-militarization-of-artificial-intelligence.

[11] James Johnson, "Rethinking Nuclear Deterrence in the Age of Artificial Intelligence," Modern War Institute at West Point, January 28, 2021, mwi.usma.edu/rethinking-nuclear-deterrence-in-the-age-of-artificial-intelligence.

[12] See Horowitz and Scharre, "AI and International Stability," Center for New American Security, January 2021; Boulanin et al., "Artificial Intelligence, Strategic Stability, and Nuclear Risk," SIPRI, June 2020; and Paul Bracken, "The Hunt for Mobile Missiles," Foreign Policy Research Institute, September 2020.

[13] Erin Dumbacher and Page Stoutland, "Nuclear Weapons Modernization: Security and Policy Implications of Integrating Digital Technology," Nuclear Threat Initiative, November 2020, media.nti.org/documents/NTI_Modernization2020_FNL-web.pdf.

[14] An AI tutorial for DoD leaders published in April 2020 by Greg Allen called "Understanding AI Technology" provides a nice overview of AI for managers and non-technical individuals interested in AI fundamentals, www.ai.mil/docs/Understanding%20AI%20Technology.pdf.

[15] John Launchbury, "DARPA Perspective on AI," www.darpa.mil/about-us/darpa-perspective-on-ai.

[16] Ed Stacey, "Artificial General Intelligence—What Does It Mean and Should We Be Worried About It?," *Forbes,* May 14, 2020, www.forbes.com/sites/edstacey/2020/05/14/artificial-general-intelligencewhat-does-it-mean-and-should-we-be-worried-about-it/#125ad66256c7.

[17] There remain skeptics about the potential for AI to continue to advance at the current rapid pace. AI has been through multiple cycles of large advancement followed by slowdowns, and many believe that the current understanding of AI remains self-limiting and an AI freeze is on the horizon. See, for example, *The Economist*, Technology Quarterly, "AI and its limits," June 11, 2020.

[18] Graphic courtesy of Marc Zissman, Massachusetts Institute of Technology Lincoln Laboratory.

[19] Michael Horowitz, Paul Scharre, and Alexander Velez-Green, "A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence," arxiv.org/abs/1912.05291.

[20] Leung, Braun, and Cuzzocrea say that "subareas of AI include robotics, computer vision, natural language processing (NLP), and machine learning." See Carson K. Leung, Peter Braun, and Alfredo Cuzzocrea, "AI-Based Sensor Information Fusion for Supporting Deep Supervised Learning," *Sensors* 19, no. 6, 2019:1345, www.ncbi.nlm.nih.gov/pmc/articles/PMC6470673.

[21] Zoubin Ghahramani. "Probabilistic machine learning and artificial intelligence," *Nature,* May 27, 2015, www.nature.com/articles/nature14541; William Fleshman, "Probability Theory: Fundamentals of Machine Learning (Part 1)," Towards Data Science (blog), January 29, 2019, towardsdatascience.com/probability-fundamentals-of-machine-learning-part-1-a156b4703e69.

[22] Joel Lehman et al., "The Surprising Creativity of Digital Evolution: A Collection of Anecdotes from the Evolutionary Computation and Artificial Life Research Communities," arXiv:1803.03453v3, 2018, arxiv.org/pdf/1803.03453.pdf.

[23] Will Knight, "Google just gave control over data center cooling to AI," August 17, 2018, www.technologyreview.com/s/611902/google-just-gave-control-over-data-center-cooling-to-an-ai, and David Silver and Demis Hassabis, "AlphaGo Zero: Starting from scratch," October 18, 2017, DeepMind blog, deepmind.com/blog/article/alphago-zero-starting-scratch.

[24] See, for example, Paul Scharre, *Army of None: Autonomous Weapons and the Future of War,* (New York: W.W. Norton & Co., 2018).

[25] Brundage et al., "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation," p. 17, arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.

[26] Szegedy et al., "Intriguing Properties of Neural Networks," 2014, arxiv.org/pdf/1312.6199.pdf.

[27] Richard Danzig, "Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority," CNAS, May 30, 2018.

[28] Michele 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," October 2020.

[29] See, for example, DARPA's Explainable Artificial Intelligence Program, www.darpa.mil/program/explainable-artificial-intelligence.

[30] Konaev, Huang, and Chahal, "Trusted Partners: Human Maching Teaming and the Future of Military AI, CSET Issue Brief, February 2021, cset.georgetown.edu/wp-content/uploads/CSET-Trusted-Partners.pdf.

[31] See, for example, lskitka.people.uic.edu/AutomationBias.pdf; and Will Heaven, *MIT Technology Review,* "Why Asking AI to Explain Itself Can Make It Worse," January 29, 2020, www.technologyreview.com/2020/01/29/304857/why-asking-an-ai-to-explain-itself-can-make-things-worse.

[32] Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," *Reuters*, October 9, 2018, www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

[33] See statement of Greg Brockman, co-founder and chief technology officer, OpenAI, before the Subcommittee on Space, Science, and Competitiveness Committee on Commerce, Science, and Transportation, United States Senate Hearing on "The Dawn of Artificial Intelligence," November, 2016. www.commerce.senate.gov/public/_cache/files/ae7e9ee3-df1b-4d94-96d1-267ebd206c48/E394EC703F89F04EB5CAEE827E16D012.mr.-greg-brockman-testimony.pdf. See also Patrick Caughill, "Microsoft Announces Its New Open Source Machine Learning Tools," Futurism, September 26, 2017, futurism.com/microsoft-announces-its-new-open-source-machine-learning-tools.

[34] "Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity," U.S. Department of Defense.

[35] Chad Steelberg, "The path to an AI-connected government," April 18, 2019, www.datacenterdynamics.com/en/opinions/path-ai-connected-government.

[36] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War,* (New York: W.W. Norton & Co., April 24, 2018).

[37] Department of Defense Directive 3000.09: Autonomy in Weapon Systems, November 21, 2012, p. 2, www.hsdl.org/?abstract&did=726163.

[38] Michael Horowitz says that "US military leaders have been very clear that they have no interest in autonomous systems, for example, armed with nuclear weapons." Horowitz in "Podcast: AI and Nuclear Weapons—Trust, Accidents, and New Risks with Paul Scharre and Mike Horowitz," Future of Life Institute, futureoflife.org/2018/09/27/podcast-ai-and-nuclear-weapons-trust-accidents-and-new-risks-with-paul-scharre-and-mike-horowitz.

[39] Jackson Barnett, "AI needs humans 'on the loop' not 'in the loop' for nuke detection, general says," fedscoop, February 14, 2020, www.fedscoop.com/ai-should-have-human-on-the-loop-not-in-the-loop-when-it-comes-to-nuke-detection-general-says.

[40] Petr Topychkanov, "Autonomy in Russian Nuclear Forces," chap. 8 in *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume I Euro-Atlantic Perspectives,* edited by Vincent Boulanin (Stockholm: Stockholm International Peace Research Institute, 2019), 68.

[41] Topychkanov, "Autonomy in Russian Nuclear Forces," 71; Hoffman, *The Dead Hand*, chap. 6; Mikhail Timoshenko, "Dead hand guarding the perimeter of Russia," February 18, 2015, tvzvezda.ru/news/krasnaya_zvezda/content/201502181414-gskc.htm (in Russian); Anton Valagin, "Guaranteed Retaliation: How the Russian Perimeter System Works," February 22, 2014, rg.ru/2014/01/22/perimetr-site.html.

[42] Franz-Stefan Gady, "Lora Saalman on How Artificial Intelligence Will Impact China's Nuclear Strategy," The Diplomat, November 7, 2018, thediplomat.com/2018/11/lora-saalman-on-how-artificial-intelligence-will-impact-chinas-nuclear-strategy.

[43] Beijing AI Principles, 2019, www.baai.ac.cn/blog/beijing-ai-principles.

[44] Michael Horowitz, Paul Scharre, and Alexander Velez-Green, "A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence," arxiv.org/abs/1912.05291.

[45] Boulanin et al., "Artificial Intelligence, Strategic Stability and Nuclear Risk," SIPRI, June 2020.

[46] Andrew Wagner, Government Matters, "Top priorities for the Joint Artificial Intelligence Center," September 15, 2019, govmatters.tv/top-priorities-for-the-joint-artificial-intelligence-center.

[47] David E. Hoffman, *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy,* (New York: Doubleday, 2009).

[48] John R. Harvey, "U.S. Nuclear Command and Control for the 21st Century," Nautilus Institute, May 24, 2019, nautilus.org/napsnet/napsnet-special-reports/u-s-nuclear-command-and-control-for-the-21st-century.

[49] Admiral Charles R. Richard, Defense Writers Group conversation transcript, January 5, 2021, cpb-us-e1.wpmucdn.com/blogs.gwu.edu/dist/2/672/files/2021/01/DWG-Admiral-Charles-R.-Richard.pdf.

[50] *Nuclear Matters Handbook 2020,* Chapter 2, "Nuclear Weapons Employment Policy, Planning, and NC3." Deputy Assistant Secretary of Defense for Nuclear Matters.

[51] United States 2018 Nuclear Posture Review, p. 58.

[52] Franz Stefan Gady, "Elsa B. Kania on Artificial Intelligence and Great Power Competition: On AI's potential, military potential, and the fallacy of an AI 'arms race,'" The Diplomat, December 31, 2019, thediplomat.com/2020/01/elsa-b-kania-on-artificial-intelligence-and-great-power-competition.

[53] Michael T. Klare, "Skynet Revisited: The Dangerous Allure of Nuclear Command Automation," Arms Control Association, April 2020.

[54] See Jeffrey Larsen, "Nuclear Command, Control, and Communications: US Country Profile," NAPSNet Special Reports, August 22, 2019, nautilus.org/napsnet/napsnet-special-reports/nuclear-command-control-and-communications-us-country-profile, and David Deptula and William LaPlante, "Modernizing US Nuclear Command, Control, and Communications," Mitchell Institute Report, MITRE Corporation, February 2019.

[55] Robert D. Critchlow, CRS Report for Congress, "Nuclear Command and Control: Current Programs and Issues," May 3, 2006, fas.org/sgp/crs/nuke/RL33408.pdf.

[56] See Anthony Barrett, "False Alarms, True Dangers? Current and Future Risks of Inadvertent U.S.-Russian Nuclear War," The RAND Corporation, p. 11, and Deptula and LaPlante, "Modernizing U.S. Nuclear Command, Control, and Communications," p. 27.

[57] For instance, MAYHEM automates the discovery and exploitation of cyber vulnerabilities so that the weaknesses can be patched. See David Brumley, "Mayhem, the Machine That Finds Software Vulnerabilities, Then Patches Them," *IEEE Spectrum*, January 29, 2019, spectrum.ieee.org/computing/software/mayhem-the-machine-that-finds-software-vulnerabilities-then-patches-them.

[58] Mikko Hypponen, "AI can be an ally in cybersecurity," February 11, 2020, venturebeat.com/2020/02/11/ai-can-be-an-ally-in-cybersecurity.

[59] Micha Musser and Ashton Garriott, "Machine Learning and Cybersecurity: Hype and Reality," Center for Security and Emerging Technology, June 2021, cset.georgetown.edu/publication/machine-learning-and-cybersecurity.

[60] Martin Giles, "AI for cybersecurity is a hot new thing—and a dangerous gamble," *MIT Technology Review,* August 11, 2018, www.technologyreview.com/s/611860/ai-for-cybersecurity-is-a-hot-new-thing-and-a-dangerous-gamble.

[61] See Kolja Brockmann, Sibylle Bauer, and Vincent Boulanin, "BIO PLUS X: Arms Control and the Convergence of Biology and Emerging Technologies," Stockholm International Peace Research Institute, March 2019, p. 14.

[62] Justin Doubleday, "Pentagon watchdog finds military's AI data and technologies at risk of cyber attack," Inside Cybersecurity, July 2, 2020, insidecybersecurity.com/daily-news/pentagon-watchdog-finds-military%E2%80%99s-ai-data-and-technologies-risk-cyber-attack.

[63] Ben Buchanan et al., "Automating Cyber Attacks: Hype and Reality," Center for Security and Emerging Technology, November 2020, cset.georgetown.edu/research/automating-cyber-attacks.

[64] "The Next Paradigm Shift: AI-Driven Cyber-Attacks," Darktrace, 2018. Darktrace, a private AI/ML company, explains, "In the future, AI-driven malware will self-propagate via a series of autonomous decisions, intelligently tailored to the parameters of the infected system. … Able to make those decisions autonomously, no C2 channel will be required for the attack to propagate and complete its mission. By eliminating the need for C2, the attack will become stealthier and more dangerous." www.darktrace.com/en/resources/wp-cyber-ai-analyst.pdf.

[65] Mikko Hypponen, "AI can be an ally in cybersecurity."

[66] Wojcocj Samek, Slawomir Stanczak, and Thomas Wiegand, "The Convergence of Machine Learning and Communications," arXiv:1708.08299.

[67] Paul Bracken, "Communication Disruption Attacks on NC3," NAPSNet Special Reports, May 28, 2020, nautilus.org/napsnet/napsnet-special-reports/communication-disruption-attacks-on-nc3/.

[68] "Basic Principles of State Policy of the Russian Federation on Nuclear Deterrence," June 8, 2020, www.mid.ru/en/foreign_policy/international_safety/disarmament/-/asset_publisher/rp0fiUBmANaH/content/id/4152094.

[69] This paper uses the definition of NC3 provided by the U.S. DoD in "Nuclear Matters Handbook 2020," chap. 2, p. 22.

[70] Fiona Cunningham, "Nuclear Command, Control, and Communications Systems of the People's Republic of China," Nautilus Institute, July 18, 2019.

[71] Dmitry Stefanovich, "Russia to Help China Develop an Early Warning System: In October, Russian President Vladimir Putin announced that Moscow was assisting Beijing in developing an early warning system," The Diplomat, October 25, 2019, thediplomat.com/2019/10/russia-to-help-china-develop-an-early-warning-system.

[72] Daniel Oberhaus, "How the US Knew Iranian Missiles Were Coming Before They Hit," Wired, January 8, 2020, www.wired.com/story/us-missile-defense-iranian-strike.

[73] Raytheon Intelligence & Space, "The Data Forge," March 16, 2021, www.defensenews.com/native/raytheon_intelligence_space/2021/03/16/the-data-forge.

[74] United States Space Force Fact Sheet, Upgraded Early Warning Radars, www.spaceforce.mil/About-Us/Fact-Sheets/Article/2197738/upgraded-early-warning-radars.

[75] Nathan Stout, "NORAD is using artificial intelligence to see the threats it used to miss," C4ISRNET, March 1, 2021, www.c4isrnet.com/artificial-intelligence/2021/03/01/norad-is-using-artificial-intelligence-to-see-the-threats-it-used-to-miss.

[76] Leonid Ryabikhin, "Russia's NC3 and Early Warning Systems," NAPSNet Special Reports, July 11, 2019, p. 7, nautilus.org/napsnet/napsnet-special-reports/russias-nc3-and-early-warning-systems.

[77] Thomas Nedwick, The Drive, The War Zone, "Take a Rare Look Inside Russia's Doomsday Ballistic Missile Warning System," February 15, 2021, www.thedrive.com/the-war-zone/39264/take-a-look-inside-russias-doomsday-ballistic-missile-warning-system.

[78] Elsa Kania, "China's Strategic Situational Awareness Capabilities: A Country Primer," Center for Strategic and International Studies, On the Radar, Spring 2019; and Fiona Cunningham, "Nuclear Command, Control, and Communications Systems of the People's Republic of China," Nautilus Institute, July 18, 2019.

[79] For example, hypersonic weapons are being developed by Russia, the United States, China, India, and others. The U.S. development of hypersonic systems is for conventional strike, but Russia has declared its Avangard boost glide hypersonic system to be nuclear capable, and China has likewise declared dual conventional and nuclear capability for its DF-ZF HGV boost-glide hypersonic system. Hypersonic boost glide systems have the characteristic of being as fast as ballistic missiles but capable of flying along the upper atmosphere to delay line-of-sight, ground-based radar detection and maneuvering during flight and on descent to make target prediction difficult. For more information, see Jill Hruby, "Russia's New Nuclear Weapon Delivery Systems: An Open Source Technical Review," NTI, November 2019.

[80] Cameron Tracy and David Wright, "Modeling the Performance of Hypersonic Boost-Glide Missiles," *Science & Global Security* 28, no. 3 (2020), scienceandglobalsecurity.org/archive/2020/12/modelling_the_performance.html.

[81] Thomas G. Roberts, "What Can 24 Satellites Do for U.S. Missile Defense?," Aerospace Security: A Project of the Center for Strategic and International Studies, October 18, 2018, aerospace.csis.org/what-can-24-satellites-do-for-u-s-missile-defense.

[82] Loren Thompson, "Raytheon Defines First Principles for Building a Space Sensor Layer," *Forbes* Aerospace and Defense, October 2, 2020, www.forbes.com/sites/lorenthompson/2020/10/02/raytheon-defines-first-principles-for-building-a-space-sensor-layer/#4e0f83e841ce; and Raymond S. Swanson and Kent R. Engebretson, "Artificial intelligence and hypersonic weapons drive sensing, fusion research," *Aerospace America 2019 Year in Review,* December 2019.

[83] Jaganath Sankaran, "A Different Use for Artificial Intelligence in Nuclear Weapons Command and Control," War on the Rocks, April 25, 2019, warontherocks.com/2019/04/a-different-use-for-artificial-intelligence-in-nuclear-weapons-command-and-control.

[84] Managing Editor, "NORAD and Defense Innovation Unit Create AI 'Pathfinder' to Detect Threats," Defense Techconnect, March 23, 2021, defensetechconnect.com/2021/03/23/norad-and-defense-innovation-unit-create-ai-pathfinder-to-detect-threats.

[85] Michael Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 1, no. 3, (May 2018): 41, doi.org/10.15781/T2639KP49.

[86] Page Stoutland, "Artificial intelligence and the modernization of US nuclear forces," chap. 7 in *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume I Euro-Atlantic Perspectives,* (Stockholm: SIPRI, May 2019), 65.

[87] Massieh Najafi et al., "Application of Machine Learning in Fault Diagnostics of Mechanical Systems," in *Proceedings of the World Congress on Engineering and Computer Science,* 2008, www.iaeng.org/publication/WCECS2008/WCECS2008_pp957-962.pdf.

[88] Phil Stewart, "Deep in the Pentagon, a Secret AI to Find Hidden Nuclear Missiles," Reuters, June 5, 2018, www.reuters.com/article/us-usa-pentagon-missiles-ai-insight/deeo-in-the-pentagon-a-secret-ai-program-to-find-nuclear-missiles-idUSKCN1J114J.

[89] Nate Frierson and Lizamaria Arias, "Artificial Intelligence Analysis Applications: A Technology Primer," Center for Strategic and International Studies (CSIS) On the Radar, July 29, 2019, ontheradar.csis.org/issue-briefs/artificial-intelligence-analysis-applications-a-technology-primer.

[90] "Submarine Detection and Monitoring: Open Source Tools and Technology," NTI, September 26, 2019, www.nti.org/analysis/articles/submarine-detection-and-monitoring-open-source-tools-and-technologies.

[91] "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, Volume II, East Asian Perspectives," edited by Lora Saalman, SIPRI, October 2019, www.sipri.org/sites/default/files/201910/the_impact_of_artificial_intelligence_on_strategic_stability_and_nuclear_risk_volume_ii.pdf.

[92] Paul Bracken, "The Hunt for Mobile Missiles: Nuclear Weapons, AI, and the New Arms Race," Foreign Policy Research Institute, September 21, 2020, www.fpri.org/article/2020/09/the-hunt-for-mobile-missiles-nuclear-weapons-ai-and-the-new-arms-race.

[93] Kris Osborn, "This Is the Pentagon's Plan to Improve Its Missile Threat System," The National Interest, June 3, 2020, nationalinterest.org/blog/buzz/pentagons-plan-improve-its-missile-threat-system-159961.

[94] See Chambers et al., "Presidential Decision Time Regarding Nuclear Weapons Employment: An Assessment and Options," Institute for Defense Analysis, June 2019. This report concludes that increasing decision time by faster early warning is likely more expensive and less effective than other approaches to increasing decision time, such as ensuring senior advisors are well prepared, prioritizing the survivability of U.S. nuclear forces, and upgrading pre-tactical warning and post-attack assessment.

[95] Chambers et al. argue that the traditional framework for U.S. decision making for nuclear weapons employment assumes that situation assessment, course of action development and evaluation, and direction of force occur linearly, but in modern reality, the three functions work in a highly iterative fashion.

[96] Andrew Imbrie and Elsa Kania, "AI Safety, Security, and Stability Among Great Powers: Options, Challenges, and Lessons Learned for Pragmatic Engagement," CSET Policy Brief, December 2019, cset.georgetown.edu/publication/ai-safety-security-and-stability-among-great-powers-options-challenges-and-lessons-learned-for-pragmatic-engagement/.

[97] TASS Russian News Agency, "AI to assist Russia military brass in decision-making, says hi-tech manufacturer," January 22, 2020, tass.com/defense/1111737.

[98] Paul McLeary, "Pentagon's Big AI Program, Maven, Already Hunts Data in Middle East, Africa," Breaking Defense, May 1, 2018, breakingdefense.com/2018/05/pentagons-big-ai-program-maven-already-hunts-data-in-middle-east-africa.

[99] Nate Frierson and Lizamaria Arias, "Artificial Intelligence Analysis Applications," CSIS On the Radar, July 20, 2019, ontheradar.csis.org/issue-briefs/artificial-intelligence-analysis-applications-a-technology-primer.

[100] Theresa Hitchens, "Roper Pushes Moving Project Maven to Air Force," Breaking Defense, June 11, 2020, breakingdefense.com/2020/06/roper-pushes-moving-project-maven-to-air-force.

[101] Harold Trinkunas, Herbert Lin, and Benjamin Loehrke, editors, "Three Tweets to Midnight: Effects of the Global Information Ecosystem on the Risk of Nuclear Conflict," Hoover Institution Press, March 15, 2020.

[102] Colin Clark, "Nuclear C3 Goes All Domain: Gen. Hyten," Breaking Defense, February 20, 2020, breakingdefense.com/2020/02/nuclear-c3-goes-all-domain-gen-hyten/.

[103] See Eliahu Niewood, Greg Grant, and Tyler Lewis, "A New Battle Command Architecture for Multi-Domain Opeations," MITRE Center for Technology and National Security, December 2019; and Frank Dimina, "Why a common data platform is the first step to JADC2," C4ISRNET, February 26, 2020.

[104] Michael T. Klare, "'Skynet' Revisited: The Dangerous Allure of Nuclear Command Automation," Arms Control Today, April 2020, www.armscontrol.org/act/2020-04/features/skynet-revisited-dangerous-allure-nuclear-command-automation.

[105] Tyler Rogoway, "Look Inside Putin's Massive New Command and Control Center," Jalopnik, November 19, 2015, jalopnik.com/look-inside-putins-massive-new-military-command-and-con-1743399678.

[106] TASS Russian News Agency, "AI to assist Russian military brass in decision-making."

[107] Elsa Kania, "China's Strategic Situational Awareness Capabilities: A Country Primer," CSIS On the Radar, Spring 2019, res.cloudinary.com/csisideaslab/image/upload/v1564246946/on-the-radar/China%20strategic%20SA.pdf.

[108] Edward Geist and Andrew J. Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?," RAND Corporation, 2018, www.rand.org/pubs/perspectives/PE296.html; *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, vol. I: Euro-Atlantic Perspectives,* edited by Vincent Boulanin, (Stockholm: Stockholm International Peace Research Institute, 2019): 18, https://www.sipri.org/sites/default/files/2019-05/sipri1905-ai-strategic-stability-nuclear-risk.pdf.

[109] Deeks, Lubell, and Murray, "Machine Learning, Artificial Intelligence, and the Use of Force," 5.

[110] Nicholas Thompson, "Inside the Apocalyptic Soviet Doomsday Machine," Wired, September 21, 2009, https://www.wired.com/2009/09/mf-deadhand/.

[111] Mikhail Tymoshenko, "Dead hand on the guard of the perimeter of Russia."

[112] Geist and Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?," p. 10; and L Ryabikhin, "Russia's NC3 and Early Warning Systems.

[113] Adm. James A. Winnefeld, Jr., "A Commonsense Policy for Avoiding a Disastrous Nuclear Decision," Carnegie Endowment for International Peace, September 10, 2019, carnegieendowment.org/2019/09/10/commonsense-policy-for-avoiding-disastrous-nuclear-decision-pub-79799.

[114] For an analysis of the "Decide Under Attack" concept, see Page Stoutland and Samantha Pitts-Kiefer, "Nuclear Weapons in the New Cyber Age," NTI, September 2018, p. 24, media.nti.org/documents/Cyber_report_finalsmall.pdf.

[115] Michael Horowitz notes that "how nuclear-armed states think about using autonomous systems may depend most on the extent to which they view their second-strike capabilities as vulnerable. The more vulnerable they view these capabilities to be, the more likely they are to integrate autonomous systems, especially those that may speed up decision-making or cut the human out of the loop." Horowitz, "Artificial intelligence and nuclear stability," in "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk," ed. Boulanin, p. 79. See also "Podcast: AI and Nuclear Weapons—Trust, Accidents, and New Risks with Paul Scharre and Mike Horowitz," Future of Life Institute, futureoflife.org/2018/09/27/podcast-ai-and-nuclear-weapons-trust-accidents-and-new-risks-with-paul-scharre-and-mike-horowitz.

[116] Technologies are also being developed to protect missile sites locally to ensure second-strike capability.

[117] Adam Lowther and Curtis McGiffin, "America Needs a 'Dead Hand,'" War on the Rocks (blog), August 16, 2019, warontherocks .com/2019/08/america-needs-a-dead-hand.

[118] Ibid.

[119] Ashley Deeks, Noam Lubell, and Daragh Murray, "Machine Learning, Artificial Intelligence, and the Use of Force by States," *Journal of National Security Law & Policy* 10, no. 1 (2019), 9–10. jnslp.com/wp-content/uploads/2019/04/Machine_Learning_Artificial_Intelligence_2.pdf.

[120] Matt Field, "Strangelove redux: US experts propose having AI control nuclear weapons," *Bulletin of the Atomic Scientists,* August 30, 2019, thebulletin.org/2019/08/strangelove-redux-us-experts-propose-having-ai-control-nuclear-weapons.

[121] Anthony Barrett, "False Alarms, True Dangers?"

[122] "Lethal Autonomous Weapons Systems," Future of Life Institute, futureoflife.org/lethal-autonomous-weapons-systems.

[123] Anton Hristozov, "The role of artificial intelligence in autonomous vehicles," Embedded Focus, July 15, 2020, www.embedded.com/the-role-of-artificial-intelligence-in-autonomous-vehicles; and "Artificial intelligence algorithms and challenges for autonomous vehicles," Embedded Focus, August 3, 2020, www.embedded.com/artificial-intelligence-algorithms-and-challenges-for-autonomous-vehicles/.

[124] Stoutland, "Artificial intelligence and the modernization of US nuclear forces," p. 65.

[125] Kris Osborn, "New Air Force B-21 stealth bomber takes key technology step toward readiness," June 2, 2020, foxnews.com, www.foxnews.com/tech/new-air-force-b-21-stealth-bomber-takes-key-technology-step-toward-war-readiness.

[126] Congressional Research Service, "Air Force B-21 Raider Long-Range Strike Bomber," updated November 13, 2019, fas.org/sgp/crs/weapons/R44463.pdf.

[127] Sebastien Roblin, "The U.S. Navy Has Orca Robot Submarines on the Way That Could Transform Naval Warfare," The National Interest, October 20, 2019.

[128] Kelley Sayler, "Artificial Intelligence and National Security."

[129] Atherton, "Is autonomy the next frontier for hypersonic vehicles?," C4ISRNET, April 29, 2019, www.c4isrnet.com/unmanned/2019/04/29/is-autonomy-the-next-frontier-for-hypersonic-vehicles.

[130] David Hambling, "U.S. Army's New Drone Swarm May Be a Weapon of Mass Destruction," *Forbes* Aerospace and Defense, June 1, 2020.

[131] Hruby, "Russia's New Nuclear Weapon Delivery Systems."

[132] Aishwarya Rakesh, "Russian Drone Attack," DefenseWorld.Net, September 5, 2020, www.defenseworld.net/feature/43/Russian_Drone_Attack#.X4ThN5NKhR4.

[133] Margarita Konaev and Samuel Bendett, "Russian AI-Enabled Combat: Coming to a City Near You?," War on the Rocks, July 31, 2019, warontherocks.com/2019/07/russian-ai-enabled-combat-coming-to-a-city-near-you.

[134] Kelley Sayler, hearing on "AI, UAVS, Hypersonics, and Autonomous Systems: Emerging Technologies and Euro-Atlantic Security," January 22, 2020, Congressional Research Service, 7-5700, www.csce.gov/sites/helsinkicommission.house.gov/files/Sayler%20-%2001222020%20Helsinki%20Commission%20Testimony%20-%20Cleared_0.pdf.

[135] For in-depth analysis of AI military use in China, refer to, for example, Elsa Kania, "Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power," Center for New American Security, November 2017, and Saalman, "The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk."

[136] Ankit Panda, "Questions About China's DF-17 and a Nuclear Capability," The Diplomat, Febuary 16, 2020, thediplomat.com/2020/02/questions-about-chinas-df-17-and-a-nuclear-capability.

[137] Elsa B. Kania, "Chinese Military Innovation in Artificial Intelligence," testimony before the U.S.-China Economic and Security Review Commission Hearing on Trade, Technology, and Military-Civil Fusion, June 7, 2019, Center for New American Security, www.cnas.org/publications/congressional-testimony/chinese-military-innovation-in-artificial-intelligence.

[138] Boulanin, ed., "Artificial Intelligence, Strategic Stability, and Nuclear Risk."

[139] Elsa Kania, "Swarms at War: Chinese Advances in Swarm Intelligence," The Jamestown Foundation, China Brief, July 6, 2017, jamestown.org/program/swarms-war-chinese-advances-swarm-intelligence.

[140] David Hambling, "This Record-Braking Shanghai Drone Display Is a Show of Technological Strength," *Forbes,* April 6, 2021, www.forbes.com/sites/davidhambling/2021/04/06/why-this-record-breaking-drone-display-in-shanghai-is-a-show-of-technological-strength/?sh=16aa6d152d53.

[141] Ahmad Y. Javaid et al., "Cyber Security Threat Analysis and Modeling of an Unmanned Aerial Vehicle System," IEEE Conference Paper, 2012, p. 586, www.researchgate.net/profile/Ahmad_Javaid/publication/235676360_Cyber_security_threat_analysis_and_modeling_of_an_unmanned_aerial_vehicle_system/links/57c6db1608ae9d64047e4cbb/Cyber-security-threat-analysis-and-modeling-of-an-unmanned-aerial-vehic.

[142] "Researchers use spoofing to 'hack' into a flying drone," BBC News, Tech section, June 29, 2012, www.bbc.com/news/technology-18643134.

[143] "Out of sight: Satellite positioning-data are vital—but the signal is surprisingly easy to disrupt," *The Economist,* July 27, 2013, www.economist.com/international/2013/07/27/out-of-sight.

[144] Jessica Cox and Heather Williams, "The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability," *The Washington Quarterly,* Spring 2021.

[145] See Robert Work, "Wargaming and Innovation," Memorandum, Deputy Secretary of Defense, February 9, 2015; and Elsa Kania, "Learning Without Fighting: New Developments in PLA Artificial Intelligence War-Gaming," The Jamestown Foundation, China Brief, April 9, 2019, jamestown.org/program/learning-without-fighting-new-developments-in-pla-artificial-intelligence-war-gaming.

[146] See Kathleen Araujo and Jose Gomera, "Disruptive Change in Unmanned Aerial Systems, Nuclear Facilities, and Radiological Protection: A Review of US and French Developments," Brookhaven National Laboratory, November 2016, BNL-113268-2016-CP; and "The Endless Aerial Surveillance of the Border," *The Atlantic*, October 11, 2019.

[147] "Nuclear Proliferation and Arms Control Monitoring, Detection, and Verification: A National Security Priority: Interim Report," National Academies of Sciences, Engineering, and Medicine, 2021, www.nationalacademies.org/our-work/review-of-capabilities-for-detection-verification-and-monitoring-of-nuclear-weapons-and-fissile-material.

[148] Geist and Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?" 6.

[149] Jason Arteburn, Erin Dumbacher, and Page Stoutland, "Signals in the Noise: Preventing Nuclear Proliferation with Machine Learning and Publicly Available Information," C4ADS and NTI report, media.nti.org/documents/Signals_in_the_Noise_-_Preventing_Nuclear_Proliferation_with_Machine_Learning__PAI.pdf.

[150] The United States has national missile defense to protect from limited strategic nuclear attacks and regional defenses to protect high-value assets. The United States assists its allies with regional missile defense systems. Russia has a long-standing ballistic missile defense system deployed around Moscow. Regional air- and missile-defense systems, the S-300 and S-400, are kinetic interceptors that are used in Russia and sold throughout Asia and Eastern Europe. Russia is developing the S-500 to defend against ballistic, cruise, and hypersonic missiles. China is developing a medium-range missile defense system and also relies on Russia's S-300 and S-400 systems for regional defense. India is also well along in developing regional missile defense systems.

[151] R. Malmathanraj and S. Mathanraj, "Solving Missile Defense and Interceptor allocation problem using Reinforcement learning and optimization techniques," *International Journal of Recent Trends in Engineering* 2, no. 3 (2009):117.

[152] Geist and Lohn, "How Might Artificial Intelligence Affect the Risk of Nuclear War?" 1.

[153] Michael Horowitz and Paul Scharre, "AI and International Stability: Risks and Confidence-Building Measures," Center for a New American Security, January 2021.

[154] See remarks by former Senator Sam Nunn at the NPT 50th Anniversary Event, www.nti.org/analysis/speeches/remarks-former-senator-sam-nunn-npt-50th-anniversary-event. "We must make technology work to reduce, not increase, risk. All nuclear-weapons states should commit to undertake a comprehensive review of their own nuclear weapons systems, and related command and control communication and warning systems, to reduce any risk of nuclear war occurring as the result of accident, miscalculation, nuclear terrorism, or cyberattack. All nuclear-weapons states should develop the technical capability to destroy any of their nuclear armed missiles if launched by mistake. We have fail-safes on satellite launches."

[155] Flournoy, Haines, and Chefitz, "Building Trust through Testing."

[156] Some concepts for fail-safe approaches to AI have been suggested; see, for example, Tobias Baumann, "An Introduction to worst-case AI safety," July 5, 2018; Trent Moore, "Google is Building a Fail Safe into AI to Shut It Down in Case It Turns Evil," SYFY Wire, June 8, 2016; and Lukas Gloor, "Suffering-focused AI safety: In favor of "fail safe" measures," Center on Long-Term Risk, Report FRI-16-1, June 2016.

[157] Paul Scharre, "Debunking the AI Arms Race Theory," Texas National Security Review, *The Strategist* 4, no. 3 (Summer 2021): 121–132, tnsr.org/2021/06/debunking-the-ai-arms-race-theory.

[158] National Security Commission on Artificial Intelligence, Final Report, March 2021, www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

[159] Jessica Cox and Heather Williams, "The Unavoidable Technology: How Artificial Intelligence Can Strengthen Nuclear Stability," *The Washington Quarterly,* Spring 2021.

About the Nuclear Threat Initiative

NTI is a nonprofit global security organization focused on reducing nuclear and biological threats imperiling humanity. www.nti.org

More from the Nuclear Threat Initiative

November 2020

U.S. Nuclear Weapons Modernization: Security and Policy Implications of Integrating Digital Technology

Erin D. Dumbacher and Page O. Stoutland, PhD.

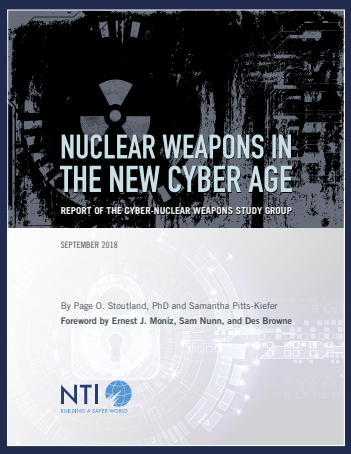
The report warns that the digitization and automation plans for the first major upgrade of the U.S. nuclear weapons systems in nearly 40 years carry significant risks and uncertainties alongside key benefits.

November 2019

Russia's New Nuclear Weapon Delivery Systems: An Open-Source Technical Review

Jill Hruby

The first detailed, exclusively open-source assessment of the five new nuclear weapon systems announced by Russian President Vladimir Putin in 2018 plus analysis on the need to extend the New START Treaty.



September 2018

Nuclear Weapons in the New Cyber Age: Report of the Cyber-Nuclear Weapons Study Group

Foreword by Ernest J. Moniz, Sam Nunn, and Des Browne

Page O. Stoutland, PhD and Samantha Pitts-Kiefer

After assessing credible, real-world scenarios, this report concludes that a successful cyberattack on nuclear weapons or related systems could have catastrophic consequences.

All papers are available at **www.nti.org**.

20 YEARS OF NTI
BUILDING A SAFER WORLD

1776 Eye Street, NW | Suite 600 | Washington, DC 20006 | **@NTI_WMD** | **www.nti.org**