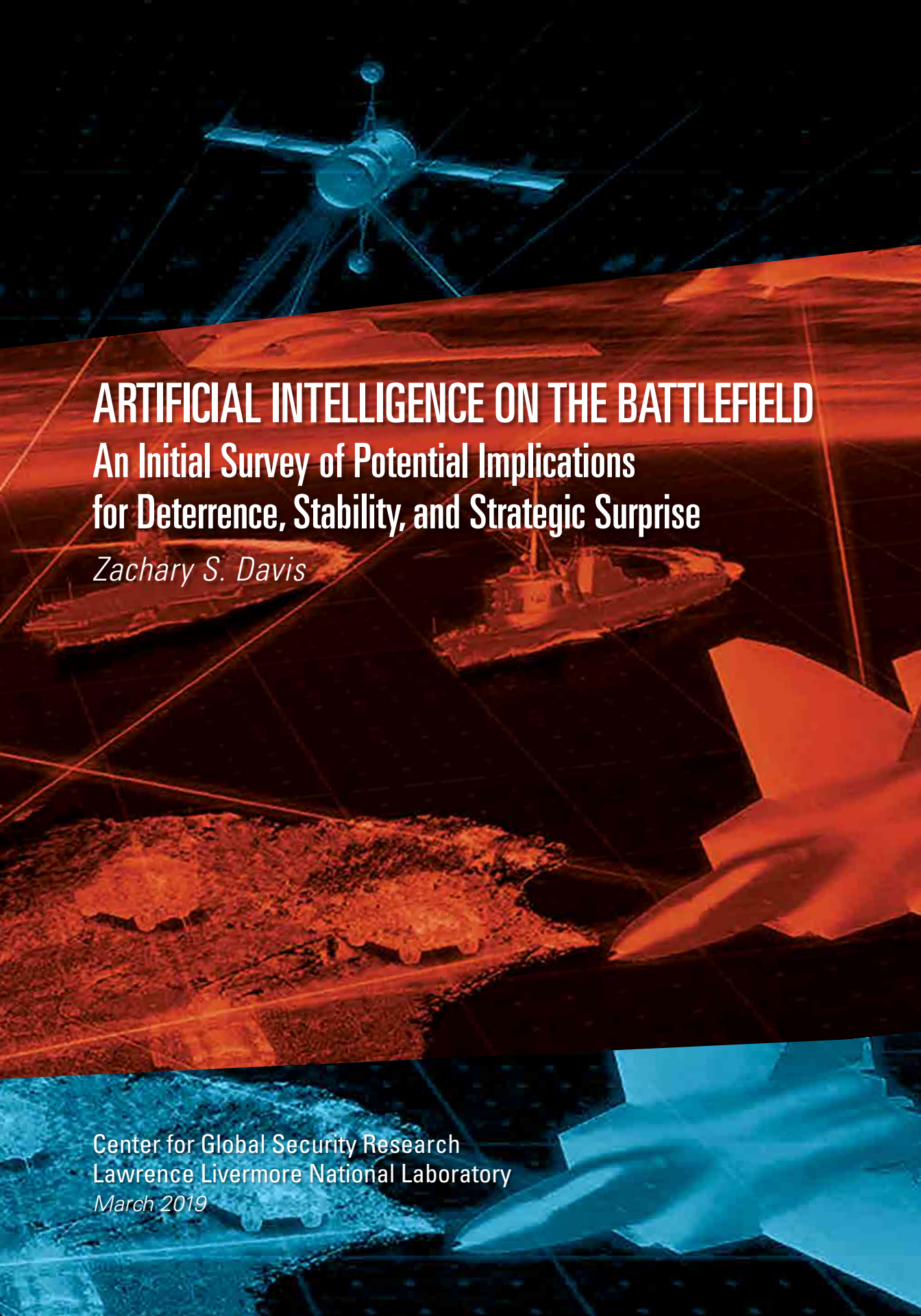


ARTIFICIAL INTELLIGENCE ON THE BATTLEFIELD

An Initial Survey of Potential Implications for Deterrence, Stability, and Strategic Surprise

Zachary S. Davis

Center for Global Security Research
Lawrence Livermore National Laboratory
March 2019

ARTIFICIAL INTELLIGENCE ON THE BATTLEFIELD

An Initial Survey of Potential Implications for Deterrence, Stability, and Strategic Surprise

Zachary S. Davis

Table of Contents

About the Author

Zachary Davis is a senior fellow at the Center for Global Security Research at Lawrence Livermore National Laboratory and a research professor at the Naval Postgraduate School in Monterey, California, where he teaches courses on counterproliferation. He has broad experience in intelligence and national-security policy and has held senior positions in the executive and legislative branches of the U.S. government. His regional focus is South Asia.

Davis began his career with the Congressional Research Service at the Library of Congress and has served the State Department, Congressional committees, and the National Security Council. Davis was the group leader for proliferation networks in LLNL's Z Program and, in 2007, senior advisor at the National Counterproliferation Center, in the office of the Director of National Intelligence. He has written many government studies and reports on technical and regional-proliferation issues and currently leads a project on the national-security implications of advanced technologies, focusing on special operations forces.

Davis's publications include articles in *Orbis, Asian Survey, Arms Control Today, Security Studies*, and *The American Interest* and chapters in edited volumes. He was editor of the widely read 1993 book *The Proliferation Puzzle: Why Nuclear Weapons Spread and What Results* and *The India–Pakistan Military Standoff,* published in 2014. He recently edited books on emerging technology: *Strategic Latency and World Power: How Technology is Changing our Concepts of Security*, and *Strategic Latency, Red, White, and Blue: Managing the National and International Security Consequences of Disruptive Technologies.*

Davis holds doctoral and master's degrees in international relations from the University of Virginia. Married to Lisa Owens Davis, he is the father of two teenaged sons and enjoys surfing and tai chi.

Introduction

Artificial intelligence has burst upon the national-security scene with a suddenness and intensity to surprise even the most veteran observers of national policy discourse. This spike of interest is driven in part by those who view AI as a revolutionary technology, on par with the discovery of fire, electricity, or nuclear weapons.[1] It is driven in part by the rapid absorption of nascent AI-based technologies into diverse sectors of the U.S. economy, often with transformative effects (as, for example, in the sciences and social media). And it is driven in part by the ambitions of America's potential adversaries. Echoing the nineteenth-century naval strategist Alfred Mahan ("Whoever rules the waves rules the world"), Russian president Putin has asserted that the nation that rules in AI "will be the ruler of the world."[2] China's president is less outspoken on this matter, but has committed China to become the dominant AI power by 2030.[3] There are mounting fears of a "Sputnik moment," which might reveal the United States to be woefully underprepared to manage new AI challenges.

What should we make of all this? Are expectations of revolutionary AI sound? Will the consequences prove positive, negative, or perhaps both for U.S. security and international stability? Definitive answers to these questions will take shape in the coming years, as we gain a better appreciation of the potential military applications of AI. At this early stage, it is useful to explore the following questions:

1. What military applications of AI are likely in the near term?
2. Of those, which are potentially consequential for the stability of strategic deterrence? Relatedly, how could AI alter the fundamental calculus of deterrence?
3. How could AI-assisted military systems affect regional stability?
4. What is the connection between regional stability and strategic deterrence?
5. What are the risks of unintended consequences and strategic surprise from AI?

This paper frames large questions and provides first-order arguments about them. It is intended to set an agenda, but not delve deeply into any particular aspect. It draws on ideas developed for a workshop convened at CGSR in September 2018 in partnership with Technology for Global Security, an NGO focused on these matters. The workshop engaged a diverse mix of public- and private-sector experts in an exploration of the emerging roles and consequences of AI. A summary of that workshop and an annotated bibliography aligned with the agenda are available at the CGSR website.[4] This paper also draws on previous work at CGSR on disruptive and latent technologies and their roles in the twenty-first-century security environment.[5]

A Realistic Appraisal of AI, Big Data, and Machine Learning

The defense community has begun its consideration of these questions with a somewhat fuzzy view of the technologies that combine to make AI. The national security community has lacked a common language for discussing AI and a detailed appreciation of the different technologies and timelines by which they might mature into militarily-significant capabilities. A more realistic appraisal can be done by assessing current trends in the scientific and business applications of AI.

The term "AI" describes a wide range of loosely related phenomena that are generally associated with using computers to glean insight from "big data." Much as the generic term "cyber" is used for everything from networks to hardware, software, automation, industrial controls, hacking, bullying, warfare, and all things social media, AI is used as a generic term that washes over meaningful distinctions among its different manifestations. This breeds confusion, especially regarding claims about its revolutionary effects.

For the vast majority of current applications, AI consists of algorithms that are the basis of pattern-recognition software. Combining this with high-performance computing power, data scientists can probe and find meaning in massive data collections. Neural networks supercharge the ability of the algorithms to identify and organize patterns in the data by "training" them to associate specific patterns with desired outcomes. Multiple layers of neural networks, known as deep-learning neural networks, make current approaches to "machine learning," "supervised learning," and "reinforcement learning" possible.[6] However, the neural-network approach covers only a fraction of the advances in AI methods. For example, AI also includes language processing, knowledge representation, and inferential reasoning, which are all increasingly possible due to advances in software, hardware, data collection, and data storage. AI represents a quantum leap in the ability to find needles in data haystacks—as long as you know what you are looking for.

It is useful to distinguish between narrow and general applications of AI. Narrow AI uses discrete problem-solving tools to perform specific narrow tasks. General AI encompasses technologies designed to mimic and recreate functions of the human brain. The gap between the two is significant. Most experts appear to agree that the accomplishments of narrow AI, though quite significant, are a long way from the requirements that must be met to replicate human-like reasoning as envisioned by proponents of general AI. Although IBM's Watson, Google's Deep Mind, and other such experiments have made breakthroughs in replicating human-like reasoning, they are a long way from reliably replicating the performance of the brain in its multiple dimensions. It is not surprising, however, that our imaginations are captivated by what futurists have called "the singularity"—a point in time

when "we will multiply our effective intelligence a billionfold by merging with the intelligence we have created."[7] The quest for "superintelligence" notwithstanding, recent progress in brain enhancement mostly replenishes impaired functioning[8] and has a long way to go before citizens, soldiers or robots are equipped with superhuman powers.[9]

Although general AI stimulates intriguing science fiction about cyborgs, space wars, and robot armies, narrow AI is already here—and has been for some time.

Narrow AI is already in broad use in the private sector. In both business and science, AI has wide applications, primarily in data-rich research fields, including fundamental research (e.g., physics, chemistry, and biology) and applied sciences (e.g., medicine, aeronautics, and environmental studies). Data science is facilitating rapid advancements in every aspect of scientific discovery, even changing long held methodological standards and practices.[10] Figure 1 illustrates some areas where AI-fueled deep learning is having its greatest effect.

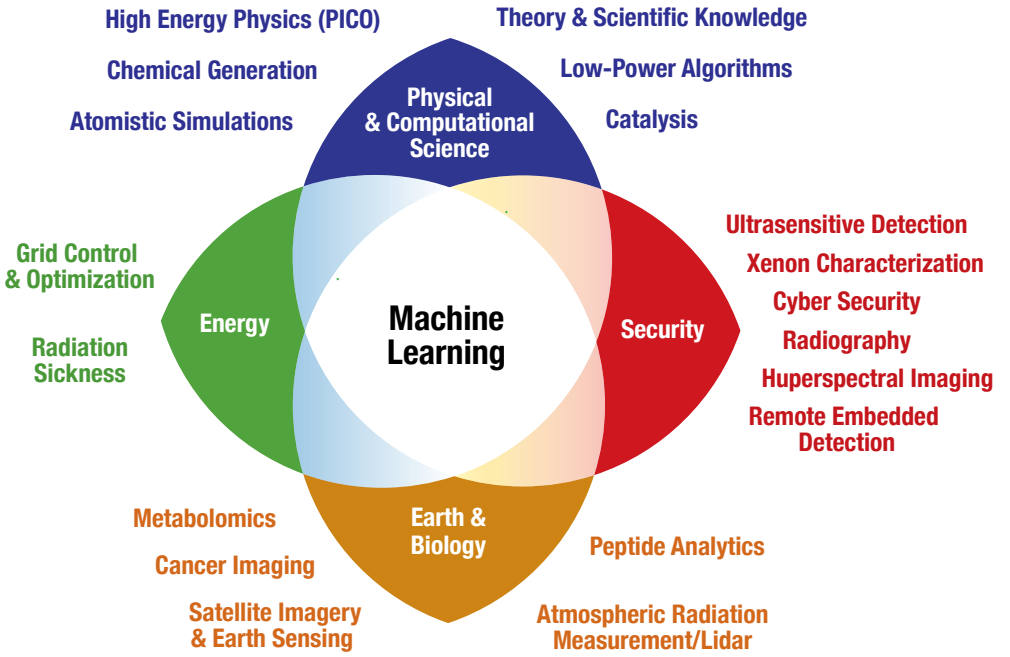


Figure 1 Disciplinary areas of deep learning for scientific discovery at the Pacific Northwest National Laboratory. SOURCE: Nathan Hodas, Pacific Northwest National Laboratory, reproduced in National Academies Press, *Artificial Intelligence and Machine Learning to Accelerate Translational Research: Proceedings of a Workshop in Brief*, July 2018, available at http://nap.edu/25197

The crossover of AI into business applications has empowered predictive analytics for market research, consumer behavior, logistics, quality control, and many other data-rich areas. The proliferation of cameras and sensors creates even more opportunities for data analysis. When combined with robotics, AI is ushering in a new industrial age, with far-reaching societal implications for labor and management.[11] For these types of applications, however, AI is more a well-established, sustaining, and enabling technology than a revolutionary new disruptive technology in its own right. Data analytics is not new, but it is getting better.

In these scientific and business applications, AI is an enabling technology, a cross-cutting force multiplier when coupled with existing data-centric systems such as the internet, heathcare, social media, industrial processes, transportation, and just about every aspect of the global economy, where recognizing patterns is the key to insight and profit. Growing interconnectivity, illustrated by the internet of things (IOT), is producing more data and providing more opportunity for AI algorithms to reveal hidden insights.

Characterizing the Military Potential of AI

Like so many technologies, AI is loaded with latent military potential.[12] How long will it be until we see game-changing AI applications in this field?

Algorithmic warfare is potentially a prime mover of a revolution in military affairs.[13] AI was central to the "third offset strategy" pursued by the Department of Defense in the second Obama term[14] and was a principal focus of multiple government initiatives to accelerate the development of advanced technologies. In June 2018, DOD set up its Joint Artificial Intelligence Center,[15] following the establishment of the White House's Select Committee on AI in May 2018[16] and the release of the White House Executive Order on Maintaining American Leadership in Artificial Intelligence on February 11, 2019. DOD and IC spending on AI has also increased.[17] For military applications with direct analogs in the civilian world, like logistics, planning, analysis, and transportation, AI-supported data analytics are already in use throughout the defense and intelligence communities.[18]

These applications are separate and distinct from applications to warfighting, which tend to fall into two categories: those with effects primarily at the operational level of war and those that also affect the strategic level. AI applications at the operational level of war could have a very significant impact on the use of general-purpose military forces to achieve tactical objectives, and thus on the credibility of conventional deterrence. AI applications at the strategic level could have significant influence on political decisions about the scale and scope of war, escalation and de-escalation, and, by extension, strategic stability and deterrence.

Illustrative AI Applications at the Operational Level of War

Three potential applications of AI at the operational level of war serve to illustrate its potentially wide-ranging implications for the use of general-purpose military forces.

***Omnipresent and Omniscient Autonomous Vehicles**:* Exploiting the new generation of autonomous vehicles is a high priority for military applications of AI, with much of the focus on navigation for a variety of unmanned land, sea, and air systems.[19] Space and undersea platforms will also benefit from AI-informed guidance systems. AI is at the heart of the so-called "drone swarms" that have received much attention in recent years.[20] AI-informed navigation software supported by ubiquitous sensors enables unmanned vehicles to find their way through hostile terrain and may eventually make it possible for complex formations of various types of drones to operate in multiple domains with complementary armaments. So equipped, they will be able to conduct sophisticated battle tactics and instantly adjust to enemy maneuvers to exploit battlefield opportunities and report changing conditions. Autonomous vehicles and robotics are poised to revolutionize warfare.

Big-Data-Driven Modeling, Simulation, and Wargaming: AI has steadily been increasing the power of simulations and gaming tools used to study nuclear and conventional weapons. From Samuel Glasstone's early calculations of nuclear effects to the extensive library of RAND studies on nuclear issues, quantitative methods have been integral to the development of nuclear weapons systems.

AI is enabling scientists to model nuclear effects to confirm the reliability of the nuclear stockpile without nuclear testing, via the stockpile-stewardship program. Simulation and modeling is already a key part of the design process for nearly all major weapons systems, from jets and ships to spacecraft and precision-guided munitions.[21] Massive modeling and simulation will be necessary to design the all-encompassing multidomain system of systems envisioned for battle management and complex missions such as designing, planning, and managing systems for space situational awareness. On the production side, AI already informs quality control for novel production methods such as additive manufacturing.[22]

AI is also enriching battlefield simulations and wargames involving multi-actor interactions. AI enables wargamers to add and modify game variables to explore how dynamic conditions (weapons, effects, allies, interventions, etc.) may affect outcomes and decision making. AI is also used to analyze the results of such games. [23] These are examples of evolutionary learning that are unlikely to cause strategic surprise or undermine stability.

Focused Intelligence Collection and Analysis: With so many incoming streams of intelligence being collected from so many sources—HUMINT, SIGINT, GEOINT, MASINT, ELINT, OSINT, etc., all requiring analysis to be useful to policy makers—the intelligence community faces the challenge of information overload.[24] This is a data-centric problem for which AI and machine learning are well suited.[25] For example, a project at LLNL uses neural networks to probe multimodal data sets (images, text, and video) in search of key indicators of proliferation activity. Machine learning also makes it possible to combine open-source trade and financial data with multiple forms of intelligence to glean insights about illicit technology transfers, proliferation networks, and the efforts of proliferators to evade detection.[26] These insights enable analysts to inform policy makers and support counterproliferation policy and actions.

Machine learning will be an important tool for all-source analysts, who are increasingly required to consider information from a combination of sources, locations, and disciplines to understand the global security environment. To the extent that better information leads to informed decisions, applying AI to these collection and analysis problems will benefit strategic stability.

Illustrative AI Applications at the Strategic Level of War

Here again, a few examples of strategic AI applications illustrate its potential.

A System of Systems Enabling Exquisite ISR: For the military, object identification is a natural starting point for AI, as it requires culling images and information collected from satellites and drones to find things of military importance, such as missiles, troops, and other relevant intelligence information. The National Geospatial Intelligence Agency (NGA) has led the charge in applying AI to military and intelligence needs.[27] But object identification is just the beginning. Intelligence, surveillance, and reconnaissance (ISR) is the key to multidomain situational awareness. This holistic awareness is increasingly critical as the battlefield extends to all domains—sea, land, air, space, and cyber—on a global scale.

Managing and making sense of the staggering amount of ISR data involved in modern warfare is a natural fit for AI—and the objective of DOD's Project Maven, also known as the Algorithmic Warfare Cross-Functional Team.[28] According to Lt. General Jack Shanahan, the director of Defense Intelligence for Warfighter Support, Project Maven was conceived as "the spark that kindles the flame front for artificial intelligence across the rest of the department."[29]

Maven's initial mission was to help locate ISIS fighters. Its implications, however, are vast. Multidomain warfare involves colossal amounts of heterogenous data

streams that can be exploited only with the help of AI. Mirroring the proliferation of sensors in the civilian world, the multidomain, hybrid-warfare battlefield has become a military version of the internet of things, teeming with vital information for assessing tactical and strategic threats and opportunities. While the ability to manage this data colossus in real time promises tremendous advantages, failure to draw meaning from that information could spell disaster.

The ability to rapidly process a flood of information from various platforms operating in multiple domains translates into two fundamental military advantages: speed and range. Moving faster than your adversary enhances offensive mobility and makes you harder to hit. Striking from farther off adds the element of surprise and minimizes exposure to enemy fire. These were central tenets of the previous revolution in military affairs that debuted in the Gulf War. AI makes it possible to analyze dynamic battlefield conditions in real time and strike quickly and optimally while minimizing risks to one's own forces. As a recent Defense Science Board study demonstrated, such integrated battle-management, command, control, communications, and intelligence (BMC3I) capabilities are well suited to finding and targeting deployed missile batteries. They may thus be the key to countering critical elements of the anti-access area denial (A2AD) strategies of Russia and China,[30] which were designed to exploit the vulnerabilities of U.S. land and sea assets in Europe and Asia. In addition to geolocating targets, AI-enabled BMC3I could help guide and coordinate kinetic effects involving multiple platforms, possibly providing a counter to current adversarial A2AD. From this perspective, the cumulative effects of tactical level AI could become a strategic-level game changer.

Precision Targeting of Strategic Assets: AI-empowered ISR that makes it possible to locate, track, and target a variety of enemy weapons systems raises the possibility of striking strategic assets, such as aircraft carriers, mobile missiles, or nuclear weapons. This capability, and perceptions about its existence, could disrupt long-held assumptions about deterrence stability, especially if it appeared possible to conduct a disarming counterforce strike against an adversary's retaliatory forces.[31] Offensive weapons that can "find, fix, and finish" a significant portion of an adversary's strategic assets,[32] combined with defensive systems that can shoot down remaining retaliatory capabilities, could challenge fundamental precepts of deterrence based on mutual vulnerability.

Effective Missile Defense: Advances in AI-enhanced targeting and navigation have improved prospects for a wide range of tactical and strategic defense systems, especially ballistic-missile defenses, by empowering target acquisition, tracking, and discrimination.[33] The convergence of powerful new offensive and defensive capabilities has rekindled fears, however, of a surprise attack that could rattle strategic stability.

AI Guided Cyber: As an inherently digital domain, the cyber realm naturally lends itself to AI applications, as illustrated by the centrality of AI algorithms in the code of social-media titans like Google and Facebook. The availability of enormous amounts of data in electronic formats is well suited to AI strengths. AI-guided probing, mapping, and hacking of computer networks can provide useful data for machine learning, including discovery of network vulnerabilities, identities, profiles, relationships, and other information that may be valuable for offensive and defense purposes.[34] Chinese applications of AI for societal surveillance purposes arouse broad concerns about the implications for privacy and democracy, while Russian influence operations have demonstrated the vulnerability of social-media platforms to manipulation.

On the offensive side, AI could locate and target particular nodes or individual accounts for collection, disruption, or disinformation. Cyberattacks on national command infrastructure and networks, for example, could be catastrophic.[35] On the defensive side of the equation, AI can help detect such intrusions and search for debilitating anomalies in civilian and military operating systems.[36] AI will equally empower offensive and defensive measures, with both positive and negative strategic effects.

In sum, AI has great potential application in the military domain, at both the operational and strategic levels of war, and may enable significant new operational and strategic advantages as the United States and others exploit these technologies.

The Negative Side of Disruptive Technologies

But the story doesn't stop there. The United States is not the only state seeking disruptive benefits. This initial survey of the potential applications and implications of AI must account also for competitive realities and the possible drawbacks of intensifying competition with adversaries.

In the classic Cold War movie *WarGames*, a young hacker breaks into a DOD supercomputer designed to use AI to plan and execute nuclear war. He engages the computer to play Global Thermonuclear War and accidentally triggers a simulated scenario of nuclear Armageddon, which is mistaken for the real thing. The computer ultimately learns that for nuclear deterrence, "the only way to win is not to play." If AI disrupts the central logic of nuclear deterrence, as understood by the nuclear powers, or fundamentally changes the precepts that support it, the strategic consequences could be far reaching—and the prospects that computers will learn to "not play" are uncertain. The following section highlights potentially destabilizing aspects of AI.

AI may be seen by others as eroding mutual strategic vulnerability, thereby increasing the risk of war. The combination of exquisite ISR with an effective defensive shield could make it tempting to conduct a disarming, decapitating, or blinding first strike at strategic targets, including nuclear command and control (NC3), early-warning radars, or dual-capable missiles and aircraft.[37] Such a revision of deterrence logic might be highly destabilizing. Shared vulnerability and assured retaliation are central concepts of mutually-assured destruction (MAD) deterrence theory. Switching the theoretical incentive from MAD to improve the odds of successfully conducting a disarming first strike could change the risk calculus that has undergirded strategic stability for decades.[38] Preventing such a revision of nuclear-deterrence logic was the essence of Vladimir Putin's claim in March 2018 that his new weapons are "invincible against all existing and prospective missile defense and counter-air-defense systems."[39] By evading perceived U.S. global strike and missile-defense capabilities, Putin's claims about new AI-guided retaliatory forces were justified as efforts to preserve MAD.

Competition to gain advantage will bring uncertainty about the future military balance. Russia, China, and other nations' advances in these same AI-enabled technologies may shift the strategic calculus as well, especially in regional contexts. For example, while Russian and Chinese A2AD systems designed to defeat U.S. regional forces may reduce allies' confidence in American security guarantees, the U.S. ability to defeat these A2AD systems with AI-accelerated ISR, BMC3I, defensive systems, and autonomous vehicles would demonstrate resolve and provide opportunities for joint U.S.–allied defense cooperation—thereby enhancing stability and deterrence. Reinforcing regional conventional deterrence is also an essential part of strategic stability.[40] However, even the perception of an imbalance that favors striking first can lead to misperception, miscalculation, and arms racing. Whatever advantages may be attained with AI are likely to evoke countermeasures that mitigate temporary unilateral advantages. Russian and Chinese interest in hypersonic vehicles and counter- space operations may fall into this category.

AI systems are vulnerable to flawed data inputs, which can cause unintended consequences. In her book *Weapons of Math Destruction*, data scientist Cathy O'Neil demonstrates how AI algorithms distort reality and lead to incorrect, misleading, and unjust decisions.[41] Perhaps the biggest obstacle to reliance on AI is the age-old problem of data reliability. AI can magnify the "garbage in, garbage out" problem.[42] Data comes from many places and is not always carefully collected or curated. Compounding the problem of faulty data and skewed results, AI often reflects human bias[43] or creates new biases based on flawed learning from the data provided.[44] Computer vision—the AI-informed object- and pattern-recognition software behind Project Maven and many other applications—is easily fooled by misleading data.[45] Differentiating between similar objects is difficult,[46] and even more challenging under

denial and deception campaigns that may, for example, use camouflage and decoys. Even when data seems accurate, AI sometimes "hallucinates" things that do not exist.[47] Transferring the inherent problems of data reliability and interpretation to the battlefield raises critical questions about the safety and reliability that accompany desirable attributes of speed and lethality. Accidentally hitting the wrong targets, for example, could have strategic consequences.

Countering many AI applications can be simple and straightforward. Adversarial manipulation of data provides many opportunities for mischief and mistakes.[48] The fact that AI is easily deceived invites efforts to sabotage its coveted military benefits.[49] By corrupting data in calculated ways, it may be possible to cause catastrophic equipment failures, miscommunication, confusion, logistical nightmares, and devastating mistakes in AI-reliant systems. The black-box nature of AI, which makes it hard to understand how and why AI makes decisions, also makes it difficult to recognize whether data is compromised and producing inaccurate outcomes, such as hitting the wrong targets or misdirecting allied forces. Data vulnerability may well be the Achilles' heel of AI.

Speedy decision making and operational execution may serve the goals of effective crisis management poorly. On October 19, 1962, only three days into the Cuban Missile Crisis, General Curtis LeMay counselled President Kennedy, "I just don't see any other solution except direct military action right now."[50] Ten days later, the crisis was resolved diplomatically. If one of the advantages of AI is rapid decision making, that same speed could be a disadvantage if it needlessly accelerates the escalation of conflict from crisis, to war, even to potential nuclear confrontation.[51] The battlefield advantages of AI-driven ISR and autonomous systems could shrink the time available for diplomats to avoid or manage crises. As currently conceived, AI-driven battlefield systems would not include real-time reporting and analysis of national and international diplomacy to avoid, control, contain, or end a conflict—violating Clausewitz's principle of war as "the continuation of politics by other means." In many cases, initial logic may dictate striking first, as General LeMay advised. Accelerated decision making could have pushed the Cuban Missile Crisis toward logical, but undesirable, outcomes. In practice, slowing things down can be the key to victory, especially when the options include nuclear weapons.

Many of the potentially positive regional deterrence effects that could eventually result from an integrated ISR, defense, and battle-management complex might not be attainable, at least not in the near term. The overarching architecture and strategy for complex, new AI-guided ISR/battle management systems does not yet exist. In fact, a proliferation of AI systems may actually complicate one of the main problems confronting U.S. military forces: effective joint operations. AI-supported weapons,

platforms, and operating systems rely on custom-built software and hardware that is specifically designed for each separate system and purpose. There is currently no master mechanism to integrate the scores of AI-powered systems operating on multiple platforms.[52] For multidomain ISR, scores of sensors, radars, weapons, and communications systems must be integrated across multiple geophysical domains. If this were not challenging enough, these systems are built and operated by different agencies, commands, and contractors, with various authorities, accesses, and procedures. Adding allies with their own AI systems to this landscape brings further complexity and risk. Without seamless integration, the hoped-for benefits of speed and lethality may prove fleeting, and the credibility of such an unproven system of systems could be called into question. Massively complex and unproven capabilities would invite problems that might be destabilizing.

Big data and machine learning may not solve the challenge of strategic warning. Designing a multiplex of AI-informed platforms that can communicate in real time requires a new generation of data fusion, integrative software, and command architectures. Pulling these pieces together to develop a holistic threat assessment that provides policy makers with strategic warning will not happen naturally. It will require herculean efforts to collect and analyze the information owned by diverse stakeholders with distinct classification systems, analytic roles, and customer loyalties. Integrating and analyzing sensitive information from diverse sources is already a challenge, especially if it must to be done quickly. While techniques such as machine learning and computer vision will help sort and prioritize the flood of intelligence information, analysts will still have to base many judgments on incomplete or unreliable information. Developing a fully integrated system capable of strategic warning will take years.

The close operation and integration of multiple AI systems, as required on the battlefield, can be expected to have unexpected results. The flip side of stovepiped systems' not talking to each other is the issue of unexpected convergences. As various AI-guided systems operate in a shared battlespace, the way separate AI-infused platforms might interact is uncertain. The unknown outcomes of friendly interactions will likely be compounded by interactions with foreign AI systems. With so much uncertainty about AI's internal, black-box mechanisms, AI-to-AI interactions are likely to produce unanticipated and inexplicable results, like choosing the wrong targets.[53] Finally, we cannot anticipate how AI will converge with technologies such as quantum computing, electromagnetic pulses, IOT, 5G, and blockchain/distributed ledgers. Potential convergences might produce strategic surprises that confuse and confound friends and foes alike, making the fog of war even more impenetrable and increasing the risk of escalation.

Whether or not there are humans in every part of decision making, the loop is getting crowded. The interface between humans and machines—where the proverbial person in the loop is supposed to exert human control—invokes critical questions about decision-making authority and organizational hierarchies.[54] Within the military, questions of rank, service branch, and responsibility for lethal actions can be contentious in the best of times, as currently seen in debates over the authorization of U.S. drone strikes.[55] Deconflicting military and intelligence missions will not be made easier. With scores of AI-informed battlefield systems operating at breakneck speed, each connected to its own chain of command, coordination among persons responsible for fast-moving battlefield operations involving a kaleidoscope of adversaries, domains, agencies, clearance levels, contractors, allies, and organizational cultures will be difficult—especially if the goal is offensive advantage via speedy decision making. Budgets, reorganizations, access, personalities, and leadership changes may have as much influence over AI capabilities as the technology itself. There will be lots of people in the loop in lots of places, each influencing how AI contributes to separate and shared objectives. Achieving desired strategic effects will require extraordinary cooperation and communication.

Public perception is a wildcard. AI algorithms are a central component of cyber-influence operations aimed at shaping public perception. It is well understood that the use and misuse of electronic media to manipulate public perceptions, including fake news, cyberbots, and deep fakes, can affect strategic stability.[56] How the public views a particular international conflict may shape leaders' decision making and build or undermine support for issues of war and peace, especially in democratic countries. Decisions to escalate conflict may be influenced by public attitudes. AI-powered tools such as cyberbots and deep-fake technology could enrage or pacify public opinion or mislead decision makers. Now that cyber conflict has become an ingrained feature of the international landscape, we should expect manipulation of public perceptions to affect crisis management, escalation, deterrence stability, and possibly nuclear decision making.

Decisions of war and peace cannot, and will not, be left to predictive analytics. There are fundamental differences in the ways data is used for scientific, economic, and logistic purposes and for predicting human behavior. Machine learning cannot reliably predict the outcomes of sports contests, elections, or international conflicts, at least within margins of error acceptable as applied to decisions involving questions of war and peace. Despite longstanding interest in predictive analytics that can warn decision makers what to expect, faith in the ability to predict incidents or outcomes of war and conflict based on big-data machine learning is fraught with misplaced optimism.[57] Much like the perils of self-driving cars, in which AI can correctly assess most, but not all, situations, a 90% success

rate in military applications could mislead decision makers and put soldiers and civilians unnecessarily at risk. All the risks stemming from unreliable (outdated, biased, compromised) data, machine-learning bias, and interpretive errors are magnified when emotions, nonrational behavior, and inherent unpredictability cloud data and decision making. The result is wider margins of error, which may be acceptable for research purposes, but not the practical and ethical demands of national security. Close is not good enough when it comes to war, especially where nuclear risks are involved.

Finally, public–private partnerships will shape the future of AI—but war remains the preserve of the state. As a quintessentially dual-use technology, AI is freely available to everyone. It is being developed and applied beyond the reach of governmental controls. As with many other dual-use technologies, governments rely on the private sector for the fundamental research and development, software, hardware, and expertise necessary for military AI use. DOD and intelligence have strong ties to Silicon Valley and have accelerated efforts to expedite acquisitions, especially for cyber products and AI.[58] Competition among nations to secure AI talent may have strategic implications, especially with respect to counterintelligence, intellectual property, and respect for international norms of behavior.

What this means in practice is that many countries will use the same experts, companies, and global supply chains to support their military AI aspirations, creating potential competitive conflicts of interest and security vulnerabilities related to the sharing of intellectual property. This dynamic is already evident in cyber markets, where Google, for example, has found it advantageous to accommodate Chinese government practices on censorship and surveillance[59] while simultaneously expressing political opposition to supporting U.S. military AI projects such as Project Maven. Global technology companies will have to weigh the costs and benefits of serving some national customers while keeping others at arm's length. The U.S. government, however, has little choice but to remain heavily dependent on the private sector to develop and implement AI strategies.[60] Such dependence could have strategic implications if it interferes with our ability to compete for top talent and cutting-edge capabilities.

AI's Potential Effects on Deterrence and Stability

With these potential strategic impacts in mind, how could AI alter the fundamental calculus of deterrence? How might the convergence of the tactically and strategically relevant factors discussed above affect the strategic balance?

First and most fundamentally, AI could erode stability by increasing the perceived risk of surprise attack. The combination of effective defenses with exquisite ISR that makes it possible to locate mobile targets and strike with speed and precision raises long-held fears of an AI-guided "bolt from the blue" first strike. While the fundamental logic of deterrence is unchanged, perceptions that an adversary has sufficient intent and capability to conduct a preemptive attack on vital assets is likely to motivate a variety of countermeasures.

Evaluating the incentive to strike first evokes consideration of Pearl Harbor, in which the US underestimated Japan's risk calculus while fully recognizing Tokyo's capacity to launch a cross-Pacific raid. AI contributions to military and intelligence capabilities do not override political considerations—with an important caveat added for the possibility of AI-fueled manipulation of public attitudes that could distort political judgment. Avoiding and deterring conflict remains a paramount responsibility for national leaders. Slightly improved odds of eliminating all but a few of an adversary's strategic weapons and shooting down any surviving retaliation with missile defenses still involves catastrophic risks—and does not even begin to answer questions about the aftermath of such a conflict.

Nevertheless, possessing the theoretical capability to conduct a disarming first strike inevitably triggers a classic security dilemma, which is guaranteed to provoke countermeasures from those threatened by enhanced striking power. Further advances in defenses against counterforce strikes would be a predictable response, as well as hardening and camouflage to evade and confuse exquisite ISR. To the extent that AI influences perceptions of intent and capability and alters the calculus of risk and reward, it will inspire new thinking about possible offensive and defensive maneuvers in the evolution of nuclear strategy.[61]

Second, AI will have a mixed impact on regional stability in Asia and Europe. Widespread deployment of AI-supported ISR platforms is likely to affect regional stability in the five- to ten-year timeframe. While the U.S. retains the lead in translating AI to the current generation of deployed platforms, China and Russia are close behind,[62] and many U.S. allies are rapidly advancing their own AI capabilities. Initially, the speed and lethality gained from AI-informed situational awareness and battle-management systems is likely to provide the U.S. and allies with options for countering Russian and Chinese A2AD. The coming architecture of ISR, BMC3I, and defensive systems appears well positioned to provide net advantages for U.S. and allied regional-security alliances. In addition to tactical military benefits, the co-

development of multidomain ISR provides opportunities for collaboration that directly address threats to allied security, especially with respect to extended deterrence relationships with allies in Asia and Europe. Strengthening regional conventional deterrence and regional extended nuclear deterrence reduces incentives for risk taking and supports broader interests in strategic deterrence. AI applications that support these objectives will have beneficial effects for strategic stability.

Third, in certain conditions, AI competition could benefit strategic stability and bolster deterrence. Global competition in military AI is already heating up. An AI arms race is underway. Whatever advantages are possible in the near term may be short lived, however, as U.S. allies, major adversaries, and many rising powers incorporate AI into their political and military strategies. In light of the rising tide that is advancing AI prospects around the world, temporary advantages are unlikely to yield lasting military predominance. For example, China and Russia will eventually have their own versions of multidomain ISR, coupled with precision strike and layered defenses. How will these capabilities influence Beijing's thinking about the U.S. role in the South China Sea, or Russian assessments of NATO's defense of the Baltics?

These are not primarily technical issues. AI is enhancing the performance of many tactical and strategic systems, but not giving definitive unilateral advantages to anyone. The nature of warfare is changing; AI is fueling many of those changes, but the fundamental calculus of deterrence remains sturdy. Competition for military capabilities that preserves a balance of power can be stabilizing.

Fourth, uncertainties about AI may bring unintended consequences for deterrence and stability. Predicting the future of technology is a risky business. We know with certainty that AI is being incorporated into a wide array of military missions with the intent of improving our knowledge of the operational environment, adversary capabilities, and the speed and precision of offensive and defensive weapons. We can usefully speculate how these developments are poised to change the face of modern warfare and how those changes might affect regional and strategic deterrence stability, based on our understanding of political and military realities. More elusive, however, is a clear picture of how AI might converge with other technologies to produce unexpected outcomes, or "unknown unknowns." Nevertheless, below are a few possibilities that could have major strategic consequences and alter the underlying realities on which regional and strategic stability are founded.

- Distorted data could lead AI systems to take unintended actions, such as incorrectly identifying and striking targets. As discussed above, data can be polluted intentionally via counter-AI methods, or occur naturally for many reasons. Unintended actions could hasten escalation and interfere with conflict management.

- Compounding the problems of distorted data, AI makes mistakes with a frequency that is untenable for decisions affecting strategic stability. Misinterpretations of data could lead to unintended actions that spark catastrophic reactions, including escalation and retaliation.

- The convergence of AI and cyber presents several possibilities for unintended consequences and strategic surprise. AI-informed cyberattacks on NC3 could present the target of such an attack with a "use it or lose it" situation, prompting early resort to nuclear weapons.

- AI supported cyber/information warfare, including fake news and deep fakes, could distort public and leadership perceptions of international events, inflaming passions and prompting escalation.

- The accelerated battle rhythm made possible by multidomain ISR could preclude diplomatic efforts to avoid or de-escalate conflict. Even if AI works perfectly to increase the speed and lethality of warfare, moving at the speed of AI might not be optimal in all cases.

- Unpredictable AI interactions with foreign and friendly platforms could produce unwanted AI calculations that misrepresent human intentions. The black box underlying AI decisions is not well understood and could produce destabilizing results, such as striking wrong targets.

- Unexpected convergences with other technologies, such as quantum computing and electromagnetic pulse, could confuse/distort offensive or defensive instructions and lead to undesirable results, such as striking wrong targets.

- If it were eventually possible through a variety of AI-supported information gathering methods, emerging technologies, and analytic tools to track strategic assets such as submarines, the sanctity of assured retaliation could come into question. Such a strategic surprise could prompt a variety of destabilizing actions, including possible movement toward launch on warning postures.

AI is Only One Piece of a Larger Puzzle

Evolutionary changes in the logic of regional and strategic deterrence are not new, nor are they necessarily harmful to U.S. national security. Efforts to integrate AI-based technologies into U.S. defense and intelligence strategies reflect the continued innovation and competitive advantages sought in support of U.S. national-security policy. Applications of AI that support U.S. nuclear forces and infrastructure, such as command and control, logistics, and stockpile stewardship, serve to reinforce strategic deterrence by bolstering the survivability and credibility of our retaliatory forces.

AI that bolsters tactical/battlefield applications can also support strategic deterrence, especially in a regional context. The connection between regional and strategic deterrence has always been important and appears even more tightly coupled as increased speed, precision, and lethality at the tactical level produce military outcomes with the potential to escalate to the strategic level. Specifically, armed conflict stemming from failure to deter Chinese or Russian aggression against U.S. regional allies may be hard to contain, especially if early battlefield victories leave one side facing humiliating defeat. The U.S. and its allies still maintain conventional superiority, and AI is likely to extend those advantages in the near term to defeat Russian and Chinese A2AD. Rather than accept defeat, however, these countries might choose an "escalate to de-escalate" strategy that includes nuclear or other unconventional weapons to mitigate the technological advantages of the U.S. and its allies, including AI supported ISR, battle management, and defenses. If military applications of AI are to advance our national-security objectives, they must be integrated with a broader strategy that reinforces deterrence at the regional and strategic levels.

Recent changes in the U.S. deterrence posture would not be undermined by advances in AI, at least in the near term. [63] However, the rapid expansion of AI military applications throughout the world warrants our highly focused attention to ensure maximum advantage for the U.S. and allies, minimize negative effects on strategic stability, and prevent strategic surprise.

Notes

1 Anthony Cuthbertson, "What's Bigger Than Fire and Electricity? Artificial Intelligence, Says Google Boss," *Newsweek*, January 22, 2018, https://www.newsweek.com/artificial-intelligence-more-profound-electricity-or-fire-says-google-boss-786531; "Elon Musk: Mark My words – AI is Far More Dangerous Than Nukes" remarks at SXSW, CNBC, March 13, 2018, https://www.cnbc.com/2018/03/13/elon-musk-at-sxsw-a-i-is-more-dangerous-than-nuclear-weapons.html; Peter Holley, "Stephen Hawking Just Got an Artificial Intelligence Upgrade, But Still Thinks AI Could Bring an End to Mankind," *Washington Post*, December 2, 2014.

2 "Whoever Leads in AI Will Rule the World: Putin to Russian Children on Knowledge Day," September 1, 2017, RT News, https://www.rt.com/news/401731-ai-rule-world-putin/ Accessed July 31, 2018.

3 Paul Mozur, "Beijing Wants A.I. to be Made in China by 2030," *The New York Times*, July 20, 2017, https://www.nytimes.com/2017/07/20/business/china-artificial-intelligence.html Accessed September 24, 2018.

4 See *Assessing the Strategic Effects of Artificial Intelligence,* Workshop Summary, Center for Global Security Research, September 2018, https://cgsr.llnl.gov.

5 See Zachary Davis and Michael Nacht, eds., *Strategic Latency: Red, White and Blue, Managing the National and International Security Consequences of Disruptive Technologies*, (Livermore, CA: Center for Global Security Research, 2018), https://cgsr.llnl.gov/research/book; Zachary Davis, Ronald Lehman and Michael Nacht, eds., *Strategic Latency and World Power: How Technology is Changing our Concepts of Security,* (Livermore CA: Center for Global Security Research, 2014) https://cgsr.llnl.gov/research/book.

6 Jürgen Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Networks*, Volume 61, January 2015, pp. 85–117, https://doi.org/10.1016/j.neunet.2014.09.003.

7 Dom Galeon and Christianna Reedy, "Kurzweil Claims That the Singularity Will Happen by 2045," *Futurism*, October 5, 2017, https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045/; "Artificial Intelligence and Life in 2030, One Hundred Year Study on Artificial Intelligence", Report of the 2015 Study Panel, Stanford University, September 2016, https://ai100.stanford.edu/sites/default/files/ai_100_report_0901fnlc_single.pdf.

8 Sara Reardon, AI-Controlled Brain Implants for Mood Disorders Tested in People, *Nature*, November 22, 2017; Antonio Regalado, "Reversing Paralysis: Scientists are Making Remarkable Progress at Using Brain Implants to Restore the Freedom of Movement that Spinal Cord Injuries Take Away," *MIT Technology Review*, https://www.technologyreview.com/s/603492/10-breakthrough-technologies-2017-reversing-paralysis.

9 Sara Reardon, The Pentagon's Gamble on Brain Implants, Bionic Limbs and Combat Exoskeletons," *Nature*, June 10, 2015; Annie Jacobsen, "Engineering Humans for War," *The Atlantic*, September 23, 2015, https://www.theatlantic.com/international/archive/2015/09/military-technology-pentagon-robots/406786/; Michael Joseph Gross, "The Pentagon's Push to Program Soldiers' Brains," *The Atlantic*, November 2018, https://www.theatlantic.com/magazine/archive/2018/11/the-pentagon-wants-to-weaponize-the-brain-what-could-go-wrong/570841.

10 David Weinberger, "Our Machines Now Have Knowledge We'll Never Understand," *Wired*, April 18, 2017, https://backchannel.com/our-machines-now-have-knowledge-well-never-understand-857a479dcc0e.

11 Darrell West, *The Future of Work: Robots, AI and Automation*, (Washington: Brookings Institution Press, 2018); Molly Kinder, "Learning to Work With Robots: AI will Change Everything. Workers Must Adapt – Or Else," *Foreign Policy*, July 11, 2018, https://foreignpolicy.com/2018/07/11/learning-to-work-with-robots-automation-ai-labor.

12 Zachary Davis and Michael Nacht, eds., *Strategic Latency: Red, White and Blue, Managing the National and International Security Consequences of Disruptive Technologies*, ibid.

13 F.G. Hoffman, "Will War's Nature Change in the Seventh Military Revolution?" Exploring War's Character and Nature, *Parameters*, 47(4) Winter 2017-18.

14 "Deputy Secretary: Third Offset Strategy Bolsters America's Military Deterrence," Department of Defense, October 31, 2016, https://dod.defense.gov/News/Article/Article/991434/deputy-secretary-third-offset-strategy-bolsters-americas-military-deterrence.

15 Memorandum from the Deputy Secretary of Defense, "Establishment of the Joint Artificial Intelligence Center," June 27, 2018.

16 *Summary of the 2018 White House Summit on Artificial Intelligence for American Industry*, The White House, May 10, 2018.

17 DARPA Announces $2 Billion Campaign to Develop Next Wave of AI Technologies, *DARPA*, September 7, 2018, https://www.darpa.mil/news-events/2018-09-07.

18 Daniel Hoadley, Nathan Lucas, *Artificial Intelligence and National Security*, Congressional Research Service, April 26, 2018; Marcus Weisgerber, "The Pentagon's New Artificial Intelligence Is Already Hunting Terrorists," *Defense One,* December 21, 2017, https://www.defenseone.com/technology/2017/12/pentagons-new-artificial-intelligence-already-hunting-terrorists/144742/; Matt Leonard, "Army Leverages Machine Learning to Predict Component Failure, *Defense Systems*, July 2, 2018, https://defensesystems.com/articles/2018/07/03/army-vehicle-predictive-maintenance.aspx.

19 National Academies of Science, *Autonomy in Land and Sea and In the Air and Space, Proceedings of a Forum*, 2018, http://nap.edu/25168.

20 National Academy of Sciences, *Counter-Unmanned Aircraft System (CUAS) Capability for Battalion and Below Operations, Abbreviated Version of a Restricted Report*, 2018, http://www.nap.edu/read/24747/chapter/1.

21 Lisa Owens Davis, "Moving at the Speed of S&T: Calibrating the Role of National Laboratories to Support National Security," in Davis and Nacht, *Strategic Latency: Red, White and Blue*, ibid.

22 "Machine Learning to Prevent Defects in Metal 3D Printed Parts in Real Time," Lawrence Livermore National Lab, *Newsline*, September 13, 2018, https://webcenter.llnl.gov/myllnl/faces/oracle/webcenter/portalapp/pages/top-story-wrapper.jspx?articleId=52535&_afrLoop=77869951013468&_afrWindowMode=0&_afrWindowId=blank#%40%3F_afrWindowId%3Dblank%26_afrLoop%3D77869951013468%26articleId%3D52535%26_afrWindowMode%3D0%26_adf.ctrl-state%3Dt66qlfya5_65.

23 Roger Smith, "The Long History of Gaming in Military Training," *Simulation & Gaming*, no. 1, 2010: Anders Drachen Christian Thurau, Julian Togelius, Georgios Yannakakis, and Christian Bauckhage, "Game Data Mining," in Magy Seif El-Nasr, Anders Drachen, Alessandro Canossa, editors, *Game Analytics: Maximizing the Value of Player Data*, (London: Springer, 2013)PONG

24 Marc Pomerleau, "Can the Intel and Defense Community Conquer Data Overload? C4ISRNET, September 5, 2018, https://www.c4isrnet.com/intel-geoint/2018/09/05/can-the-intel-and-defense-community-conquer-data-overload/?utm_source=Sailthru&utm_medium=email&utm_campaign=daily%20brief%209/5/18&utm_term=Editorial%20-%20Daily%20Brief.

25 Marc Pomerleau, "Here's How Intelligence Agencies Will Take Advantage of Machine Learning and AI," *C4ISRNET*, May 1, 2018, https://www.c4isrnet.com/intel-geoint/2018/05/01/heres-how-intelligence-will-take-advantage-of-machine-learning-and-ai.

26 "Deep Learning to Advance Nuclear Nonproliferation," *LLNL Newsline*, August 21, 2018, https://webcenter.llnl.gov/myllnl/faces/oracle/webcenter/portalapp/pages/top-story-wrapper.jspx?articleId=52206&_afrLoop=17092740707130&_afrWindowMode=0&_afrWindowId=null#%40%3F_afrWindowId%3Dnull%26_afrLoop%3D17092740707130%26articleId%3D52206%26_afrWindowMode%3D0%26_adf.ctrl-state%3D1af49b8608_69.

27 Ben Conklin, "How Artificial Intelligence is Transforming GEOINT," *GCN*, April 18, 2018, https://gcn.com/articles/2018/04/18/ai-transform-geoint.aspx; Sandra Erwin, "NGA official: Artificial Intelligence is Changing Everything, We Need a Different Mentality," *Spacenews*, May 13, 2018, https://spacenews.com/nga-official-artificial-intelligence-is-changing-everything-we-need-a-different-mentality.

28 Kelsey Atherton, "Targeting the Future of the DOD's Controversial Project Maven Initiative," *C4ISRNET*, July 27, 2018, https://www.c4isrnet.com/it-networks/2018/07/27/targeting-the-future-of-the-dods-controversial-project-maven-initiative.

29 Jack Corrigan, "Project Maven Uses Machine Learning to Go Through Drone Video Feeds, But That Just the Beginning, Air Force Lt. Gen Shanahan Said," *Nextgov*, November 2, 2017, https://www.nextgov.com/cio-briefing/2017/11/three-star-general-wants-artificial-intelligence-every-new-weapon-system/142225.

30 Defense Science Board, "Study on Countering Anti-Access Systems with Longer Range and Standoff Capabilities: Assault Breaker II," 2017 *Summer Study on Long Range Effects,* Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, June 2018.

31 Edward Geist and Andrew Lohn, *How Might Artificial Intelligence Affect the Risk of Nuclear War? RAND*, 2018; Paul Bracken, "The Intersection of Cyber and Nuclear War," *The Strategy Bridge,* January 17, 2017, https://thestrategybridge.org/the-bridge/2017/1/17/the-intersection-of-cyber-and-nuclear-war.

32 Jeremy Hsu, "AI Can Help Hunt Down Missile Sites in China," *Wired*, November 21, 2017, https://www.wired.com/story/ai-can-help-hunt-down-missile-sites-in-china.

33 Jen Judson, "Hyten: To Address Russian and Chinese Missile Threats, It's All About the Sensors, " *Defense News,* August 7, 2018, https://www.defensenews.com/digital-show-dailies/smd/2018/08/07/hyten-to-address-russian-and-chinese-missile-threats-its-all-about-the-sensors.

34 Jack Corrigan, "DARPA Wants to Find Botnets Before They Attack," *Defense One*, September 12, 2018, https://www.defenseone.com/technology/2018/09/darpa-wants-find-botnets-they-attack/151199/?oref=defenseone_today_nl.

35 "Nuclear Weapons in the New CyberAge: A Report of the Cyber-Nuclear Weapons Study Group," Nuclear Threat Initiative, September 2018, https://www.nti.org/media/documents/Cyber_report_finalsmall.pdf.

36 Michael Sulmeyer and Kathryn Dura, "Beyond Killer Robots: How Artificial Intelligence Can Improve Resilience in Cyber Space," Sept 6, 2018, *War on the Rocks*, https://warontherocks.com/2018/09/beyond-killer-robots-how-artificial-intelligence-can-improve-resilience-in-cyber-space.

37 James Acton, "Escalation through Entanglement: How the Vulnerability of Command and Control Systems Raises the Risks of an Inadvertent Nuclear War," *International Security*, Volume 43, Summer 2018.

38 Kier Lieber and Daryl Press, "The New Era of Counterforce: Technological Change and the Future of Nuclear Deterrence," *International Security,* Volume 41, Issue 4, Spring 2017.

39 August Cole and Amir Husain, "Putin Says Russia's New Weapons Can't Be Beat. With AI and Robotics, They Can," *Defense One*, March 13, 2018, https://www.defenseone.com/ideas/2018/03/putin-says-russias-new-weapons-cant-be-beat-ai-and-robotics-they-can/146631.

40 Dave Johnson, *Russia's Conventional Precision Strike Capabilities, Regional Crises, and Nuclear Thresholds*, Livermore Papers on Global Security, No. 3, February 2018; John Warden, *Limited Nuclear War: The 21st Century Challenge for the United States*, Livermore Papers on Global Security, No.4, July 2018. Available at: cgsr.llnl.gov.

41 Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, (New York: Broadway Books, 2017).

42 Hillary Sanders, Joshua Saxe, "Garbage In, Garbage Out: How Purportedly Great Machine Language Models Can Be Screwed Up by Bad Data," *Proceedings of Blackhat 2017*, July 2017, Las Vegas, NV.

43 Jesse Emspak, "How a Machine Learns Prejudice," *Scientific American*, December 29, 2016, https://www.scientificamerican.com/article/how-a-machine-learns-prejudice.

44 ProPublica, "Machine Bias," May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; Will Knight, "Forget Killer Robots, Bias Is the Real AI Danger," *Technology Review*, October 3, 2017, https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger.

45 Louise Matsakis, "Researchers Fooled a Google AI Into Thinking a Rifle Was A Helicopter," *Wired*, December 20, 2017, https://www.wired.com/story/researcher-fooled-a-google-ai-into-thinking-a-rifle-was-a-helicopter.

46 Daniel Cebul, "Differentiating a Port from a Shipyard is a New Kind of Problem for AI," *C4ISRNET*, September 18, 2018, https://www.c4isrnet.com/intel-geoint/2018/09/18/differentiating-a-port-from-a-shipyard-is-a-new-kind-of-problem-for-ai/?utm_source=Sailthru&utm_medium=email&utm_campaign=Daily%209/19&utm_term=Editorial%20-%20Daily%20Brief.

47 Anna Rohrbach, Lisa Anne Hendricks, Kaylee Burns, Trevor Darrell, Kate Saenko, "Object Hallucination in Image Captioning," *Cornell University Library*, https://arxiv.org/abs/1809.02156.

48 Sandia National Laboratory, *Counter Adversarial Data Analytics*, SAND2015-3711, May 8, 2015.

49 Defense Science Board, Memorandum for Chairman, "Terms of Reference," Defense Science Board Task Force on Counter Autonomy, June 18, 2018, https://www.acq.osd.mil/dsb/TORs/2018_TOR_CounterAutonomy_18Jun2018.pdf.

50 Tim Weiner, "Word for Word, The Cuban Missile Crisis: When Kennedy Faced Armageddon and His Own Scornful General," *New York Times*, October 5, 1997, https://www.nytimes.com/1997/10/05/weekinreview/word-for-word-cuban-missile-crisis-when-kennedy-faced-armageddon-his-own.html.

51 Paul Scharre, "A Million Mistakes a Second," *Foreign Policy*, September 12, 2018, https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war.

52 Lawrence Livermore National Laboratory, "Building a Network of Collaborative Autonomous Machines," *Science and Technology Review*, June 2018; Mark Pomerleau, "To Win Future Conflicts, Combatant Commands Must be Integrated," *C4ISRNET*, August 15, 2018, https://www.c4isrnet.com/show-reporter/dodiis/2018/08/14/to-win-future-conflicts-combatant-commands-must-be-integrated/?utm_source=Sailthru&utm_medium=email&utm_campaign=Daily%208/15&utm_term=Editorial%20-%20Daily%20Brief.

53 Will Knight, "The Dark Secret at the Heart of AI," *Technology Review*, April 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai.

54 Michael Piellusch and Tom Galvin, "Is the Chain of Command Still Meaningful?" *War Room*, US Army War College, September 6, 2018, https://warroom.armywarcollege.edu/articles/chain-of-command.

55 Stimson Center, "*An Action Plan on US Drone Policy: Recommendations for the Trump Administration*," 2018, https://www.stimson.org/sites/default/files/file-attachments/Stimson%20Action%20Plan%20on%20US%20Drone%20Policy.pdf.

56 Herb Lin, "Developing Responses to Cyber-Enabled Information Warfare and Influence Operations," *Lawfare*, September 6, 2018, https://www.lawfareblog.com/developing-responses-cyber-enabled-information-warfare-and-influence-operations.

57 Kori Schake, "Why We Get It Wrong: Reflecting on the Future of War," Book Review of Lawrence Freedman, *The Future of War: A History, War on the Rocks*, August 10, 2018, https://warontherocks.com/2018/08/why-we-get-it-wrong-reflections-on-predicting-the-future-of-war/; Richard Danzig, *Driving in the Dark: Ten Propositions About Prediction and National Security*, Center for a New American Security, October 2011.

58 Frank Gac, Timothy Grayson, Joseph Keogh, "What Works? Public-Private Partnerships for Development of National Security Technology," in Davis and Nacht, eds, *Strategic Latency Red, White and Blue*, ibid.

59 Suzanne Nossel, "Google is Handing the Future of the Internet to China," *Foreign Policy*, September 10, 2018, https://foreignpolicy.com/2018/09/10/google-is-handing-the-future-of-the-internet-to-china.

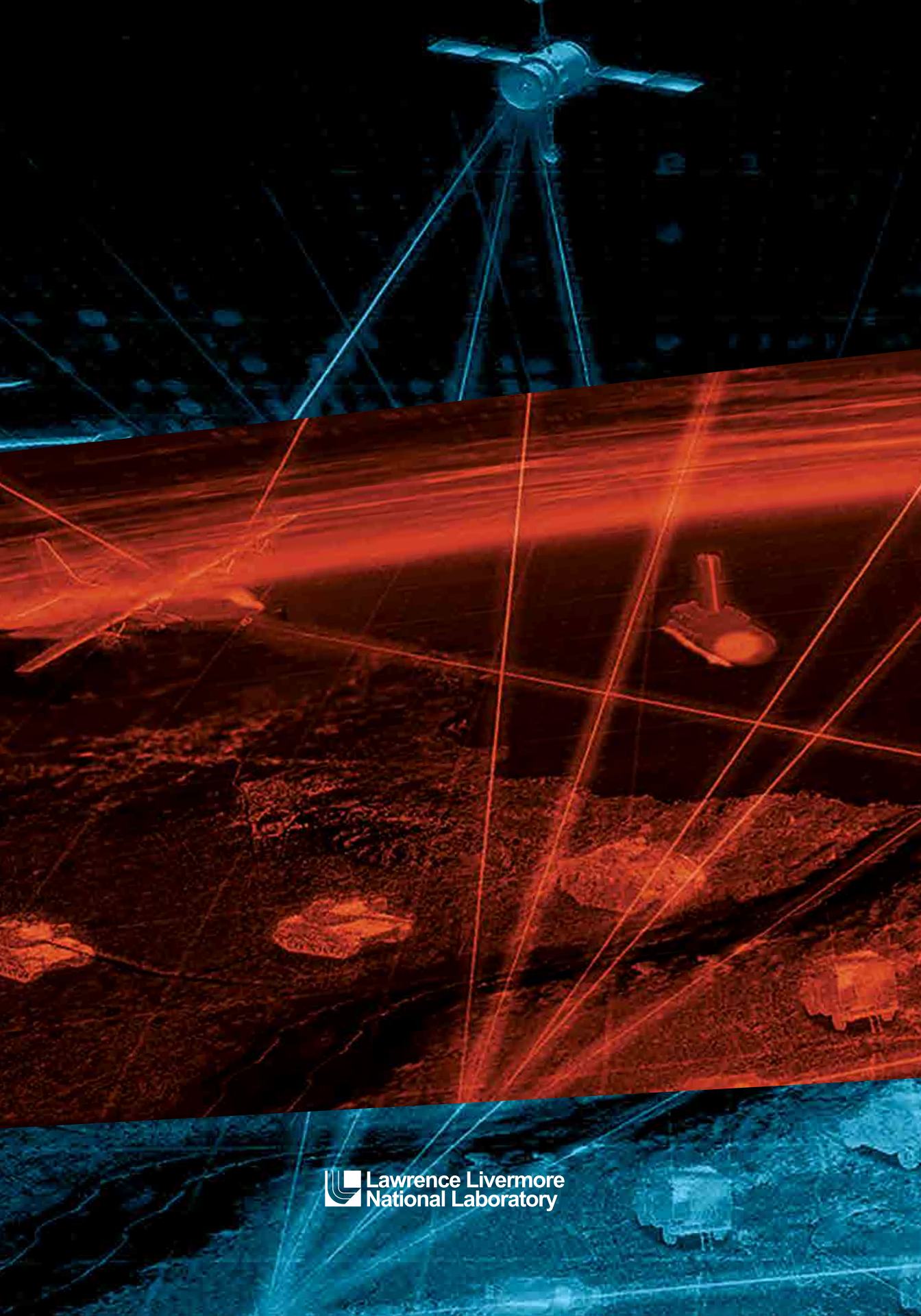
60 Laura Seligman, "Why the Military Must Learn to Love Silicon Valley," September 12, 2018, *Foreign Policy*, https://foreignpolicy.com/2018/09/12/why-the-military-must-learn-to-love-silicon-valley-pentagon-google-amazon.

61 Lawrence Freedman, *The Evolution of Nuclear Strategy*, (New York: St. Martin's Press, 1981).

62 Elsa Kania and John Costello, "*Quantum Hegemony: China's Ambitions and the Challenge to US Innovation Leadership*," Center for a New American Security, September 12, 2018, https://www.cnas.org/publications/reports/quantum-hegemony.

63 Department of Defense, "2018 Nuclear Posture Review," https://dod.defense.gov/News/SpecialReports/2018NuclearPostureReview.aspx.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC. LLNL-TR-767528 TID-19-52604


Lawrence Livermore
National Laboratory