Chapter 11

The Perfect Sentinel: AI Military Applications and Implications

By James E. Baker

AI technology is likely to be a transformative military technology, on a par with the invention of aircraft and nuclear weapons.

— Belfer/IARPA 2017 Study

It comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world.

— Vladimir Putin

…the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow.

— An Open Letter from AI & Robotics Researchers

IJCAI Conference July 28, 2015


ARTIFICIAL INTELLIGENCE AND THE FUTURE OF WARFARE

I. INTRODUCTION

Artificial intelligence (AI) is sure to influence U.S. national security policy and law, including in the areas of intelligence, homeland security, and decision-making. Nowhere is it more likely to transform security than in the area of military planning, operations, and weapons design and employment. A 2017 Belfer Center Study prepared for the Intelligence Advanced Research Projects Activity (IARPA) concluded that AI is likely to be as transformative a military technology as aviation and nuclear weapons were before. The Department of Defense thinks so. The DOD has made AI a centerpiece of its Third-Offset Strategy. In June 2018, bureaucracy followed concept as the DOD established a Joint Artificial Intelligence Center (JAIC) to facilitate and coordinate the integration of AI across DOD. This chapter considers why and how AI might transform military operations. The chapter that follows considers how the law, including the law of armed conflict applies, or might apply, to AI-enabled operating and weapons systems.

One reason AI is transformative is because depending on how it is defined the AI rubric can reach almost any or every technological innovation involving hardware, software, algorithms, big data, the IoT, autonomy, machine learning, natural language processing, data visualization, and so on. Thus, when China announced in July 2017 that it would become the world’s leader in AI by 2030 and spend $150 billion doing so, one would be hard pressed to identify precisely where and how the money will be spent. What is not opaque is China’s commitment to AI as a security and economic tool as well as a means of social control. AI will transform the security landscape. An Open Letter from AI researchers at the 2015 International Joint Conference on Artificial Intelligence Conference called autonomous weapons the Kalashnikovs of tomorrow.

At the same time, some scholars have noted that there is a lot that AI cannot currently do, like infer knowledge it does not already possess or express intuition or judgments about new information. Chatham House scholar M.L. Cummings writes, “Expertise leverages judgment and intuition as well as the quick assessment of a situation, especially in a time-critical environment such as weapons release… In humans, the ability to cope with the highest situations of uncertainty is one of the hallmarks of the true expert, but in comparison such behavior is very difficult for computers to replicate.”

One question is: When will tomorrow come? In a 2016 survey conducted by scholars at the Future of Humanity Institute, AI Impact, and Yale AI researchers from academia, industry, and government were asked when AI might exceed human performance. The median answer for respondents from China was 28 years. The median response for American specialists was 76 years.

A second question is: Will we be ready when the AI arms revolution comes?

A. Roadmap

This chapter considers the military applications and implications of AI. AI holds great promise—transformative promise—to revolutionize the design and employment of weapons. It is also certain to transform the intelligence function in tactical military contexts as well as national strategic contexts. Both developments in turn will transform the speed and content of military decision-making, logistics, and training. These changes do not come without risk. This chapter identifies four such risks:

1. The speed and volume of information as well as decision points that AI enables and thus creates, can overwhelm and undermine the capacity of military actors to make sound and ethical decisions;
2. The risk of unintended consequences, including AI that does not work as intended or exceeds or confuses the capacity of human’s to interface with technology;
3. The influence of AI on foreign relations and authoritarian regimes and the corresponding impact on international stability; and,
4. The advent and impact of a technological arms race and the corresponding influence on safety, security, first-use doctrine along with the opportunity costs that come with arms races.

Chapter 23 addresses the U.S. policy response, largely at the Department of Defense, followed by the identification of pending legal questions. The first set of questions are addressed to the defense industrial base and the tools currently available to the government to harness industrial and academic AI research and development for national security purposes. The section reviews existing authority as well as identifies pending issues, involving the scope of authority as well as potential gaps and ambiguities. One question is whether the private AI corporate community shares a sense of obligation or mutual interest with the military that traditional defense contractors once shared, and if not, whether such a relationship should be regulated by law or depend entirely on market mechanisms and litigation. Chapter 23 also identifies five law of armed conflict principles applicable to the development, deployment and use of AI-enabled systems and weapons, including the requirement to subject new weapons to legal review, the requirement to train personnel in the law, and command responsibility. This subsection asks: How do and should these principles apply to AI? Are there gaps? And should those gaps be filled with new law and, potentially, prohibitions?

B. Caveats

Chapters 11 and 23 are written from a U.S. national security policy and law perspective. AI has the potential, already realized, to assert social control in countries governed by authoritarian regimes, such as China’s Social Credit policy. It can also magnify and compound questions of privacy and security manifest in the use of social and other media. The focus of these chapters, however, is on U.S. military applica-
tions. Further, these chapters are based on and derived exclusively from unclassified sources. Further, they do not necessarily reflect the views of any branch, agency, or entity of the U.S. government. The views are those of the author alone, and thus so too are any mistakes.

II. MILITARY APPLICATIONS

There are no shortages of current and potential military applications for AI. However, the focus of AI military legal and doctrinal discussion has been on lethal autonomous weapons (LAWs). However, AI in military context offers much more. This becomes evident if one defines AI as the Stanford 100 Year Study did, as a set of technologies. “Artificial intelligence is a science, and a set of computational technologies, that are inspired, but typically operate quite differently from, the way people use their nervous systems and bodies to sense, learn, reason, and take action.”5 This definition is useful because it captures the breadth of AI-related research, its reliance on computational capacity (and prediction), and deters the reader from anthropomorphic identification with AI-enabled machines. Indeed, the operative descriptive phrase for AI capacity is Human Level Machine Intelligence (HLMI), not Human Like Machine Intelligence. AI has also been defined as everything that might happen tomorrow; yesterday’s AI now being today’s generally applied software application.

Commentators generally divide AI into two categories, narrow AI and strong AI. (Others sometimes use the terms AI and artificial general intelligence to describe these concepts.) Narrow AI is AI that can perform a single task at human levels, or perhaps better. Thus, a better term for narrow AI would be task specific AI. And, a simple definition for AI might be a field of computational capacities optimizing the ability of machines to perform programmed tasks. Narrow AI today is generally good at, and sometimes better than humans (BTHLMI), at the following tasks, usually derivative of the capacity to discern and recognize patterns and engage in statistical correlations:

- Classifying
- Predicting
- Recommending
- Translating

This has prompted DOD to identify a number of areas where AI “has massive potential” including command and communications (C2), navigation, perception, obstacle detection, and swarm behavior and tactics.\(^6\)

Strong AI, is AI that can perform multiple tasks and can switch from task to task and know when to do so, without additional human intervention (beyond that of creating the AI software and hardware in the first place).

Defined in this manner, AI becomes a military force multiplier in at least five often interlocking ways.

First, it can enable machines, shaped or not-shaped to look like animate objects—birds, dolphins, and robots, to perform inherently dangerous tasks. Depending on the capability and the task, it may perform these tasks better than humans, and certainly more safely (to the operating humans), like bomb detection and disposal, and CBRNE detection. (In other words, helping to control for danger, risk, and fear.)

Second, as an AI-enabled economy may automate and eliminate many repetitive tasks currently performed by humans, AI may have the same effect on military personnel needs and demands. Theater and garrison logistics is a case in point. AI-enabled machines, for example, may eliminate or reduce the need for personnel to provide meal and laundry services, including military and contract personnel, along with many of the costs associated with these tasks. (In other words, engaging in the sort of repetitive tasks that are essential to sustain an army, but that drain manpower.) However, as discussed in the section describing challenges, which follows, unmanned AI-enabled platforms often require a lot of manpower to operate.

Third, AI-optimized machines are less prone to human types of error brought on by fear or fatigue.

Fourth, AI-enabled systems can perform tasks not only faster than humans, but instantly, based on their capacity to compute, sort, and structure. Thus, AI-enabled systems can plan logistics supply routes and transportation schedules faster than humans.

Fifth, AI in many cases, is already better than humans at sorting vast amounts of information, characterizing that information, linking that information, and making predictions based on that information. In other words, it can bring to the military decision-maker instantaneous sources of intelligence and intelligence analysis, while also spotting anomalies and patterns predictive of risk or attack.

If you want to know how militaries might use AI beyond weapons systems, consider how an AI-enabled machine might mitigate five of the factors identified above—risk, repetition, fear, fatigue, and speed—if a machine could be “taught” to perform the task in question. That is AI. And here, the virtue or benefit derives not from the ability of the machine to act like or with human intelligence, but expressly from the fact that the machine is not human. In this sense, AI is the perfect sentinel or wingman, one that does not fall asleep on post, talk, smoke, or show fear. AI is the sentry that actually observes the Second General Order issued to sentries: “To walk

my post in a military manner, keeping always alert, and observing everything that takes place within sight or hearing.”

With that introduction, the impact of AI on military operations can be divided into four areas: Weapons, intelligence, administration and logistics, and decision-making.

**A. Weapons**

Commentators have focused on the potential of AI to enable autonomous weapons systems (AWS), including ones that look like robots, and especially ones that can be used as Lethal AWS (LAWS). The Russian military, for example, is testing a robot named FEDOR, which can fire weapons and carry heavy loads. Defense Department vernacular generally refers to AWS, but also with almost as much frequency RAS, robotic and autonomous systems. Of course, much of the research is intended to create AI systems intended to make existing weapons platforms better, like fire support systems for tanks and electronic warfare modules for aircraft.

Militaries have had some form of AWS for years. In the U.S. military inventory these systems include the Aegis ship defense system and the Counter-Rocket, Artillery, and Mortar (C-RAM) system. A heat-seeking (infrared homing) missile may also be viewed as an autonomous weapon once fired. Such a missile, the Sidewinder air-to-air missile, has been deployed with U.S. aircraft since 1956. So what is new? What is transformative?

A number of militaries, including the U.S. military, are experimenting with AI-enabled “swarms.” This is not secret. 60 Minutes did a segment titled “The Coming Swarm” in August 2017. What is secret is the trajectory of progress, capacity, date of deployment, and potential doctrinal uses. Swarms can be composed of unmanned formations of aerial, vehicular, maritime, or submarine platforms, or in current vernacular “unmanned aerial vehicles” (UAV), “unmanned ground vehicles” (UGV), “unmanned maritime vehicles” (UMV), and “unmanned underwater vehicles” (UUV). Swarms thus illustrate how robotics, autonomy, and AI work together in interlocking ways to create new capabilities. Swarms can be programmed to work in coordination with, or independent of, the command of human operators as well as manned vehicles. Imagine chaff fired from the side of an aircraft designed to fool incoming missiles. But this chaff is not comprised of metal fragments. It consists of AI-enabled pods that can maneuver around one or more incoming missiles constantly adjusting speed and trajectory, before deciding whether to attack the missile(s), lead them into the ground, or perhaps direct them back to their point of origin. This is all based on AI, because the system depends on calculations and adjustments too complex and too rapid for a human to make in the moment, let alone a human sitting in the cockpit of an aircraft contending with the pressures of combat.

Now flip from defense to offense and imagine the capability of a swarm—hundreds perhaps thousands of flying objects, called birds, projectiles, or robots—that can be used to attack naval vessels or airfields, like Kamikazes at Okinawa, but in sync and without the moral and supply chain challenges of recruiting and expending pilots. It should be no surprise then that the militaries of multiple states are working on swarm technology.

Swarms do not present the only offensive and defensive military uses for AI. Defense Department literature describes AI-enabled systems as the perfect wingman. Unmanned Systems with integrated AI, acting as a wingman or teammate, with lethal armament could perform the vast majority of the actions associated with target identification, tracking, threat prioritization, and post-attack assessment while tracking the position and ensuring the safety of blue-force assets—minimizing the risk to its human teammates.9

Further, it goes without saying that whatever is occurring in the cyber domain will be “enhanced” by AI. Operators have described hand-to-hand combat in cyberspace between humans. AI can be used, if it is not already in use, to both enable cyber combat as well as to cloak efforts at cyber attribution. If AI can be used to run stock market trading platforms that automatically buy and sell on the basis of finite differences in price, calculated to fractions of a cent, AI can surely be used to increase the speed and sophistication of cyber offense and defense. It can also be used to spoof such events.

Here definitions are important, as are distinctions between what is autonomous, what is automated, and what is used to augment human capacity. In current DOD vernacular:

An automated system is one that automatically responds or acts without human decision or input.

“Autonomy is defined as the ability of any entity to independently develop and select among different courses of action to achieve goals based on the entity’s knowledge and understanding of the world, itself, and the situation.”

An autonomous system is one that can operate on its own, but not necessarily without human input or direction. Autonomous systems are governed by broad rules that allow the system to deviate from the baseline. The Phalanx is such a system.

Augmentation, in turn, is the process by which a human and autonomous system work together, with the autonomous system augmenting human capacity. The Aegis air defense system is an example. The concept is not new, the capacity is. This is referred to by some as “the Centaur Model.”10

In short, AI can enable weapons and weapons systems to identify and engage targets in all domains rapidly, continuously, simultaneously, and sequentially and do so in a manner humans could not do, or would take too long to do in order to calculate distances, angles, numbers, and response choices. A 2018 report written by the U.N. Institute for Disarmament Research makes an additional and important distinction. “Intelligence is a system’s ability to determine the best course of action to achieve its goals. Autonomy is the freedom a system has in accomplishing its goals.”

One question is whether such systems will be, or should be, empowered to do so autonomously without affirmative human activation, choice, and decision, but based on programming and sensors alone. The United States has taken the position that with lethal autonomous systems “We will always have a human being in the loop.”

“In the loop” generally means that an autonomous system is programmed or designed to only perform its task upon human direction or command, which usually refers to the decision to fire a weapon or use force. Human “on the loop” generally refers to a system where a human is supervising the machine’s use, e.g., targeting process, and can intervene at any time during the cycle. Human “out of the loop” means the system is free to operate, e.g., select and engage targets, without subsequent human decision, supervision, or intervention. Of course, the terms are a bit deceiving, as a human is involved in writing the code and designing the system whether it is ultimately described as one with a human in, on, or out of the loop. Moreover, what it actually means to “supervise” a system on the loop may vary widely depending on the nature of engagement.

According to the 2018 DOD Roadmap Report, “DoD does not currently have an autonomous weapon system that can search for, identify, track, select, and engage targets independent of a human operator’s input.” Of course, this is a pregnant sentence, as one does not know whether the caveat derives from a singular verb (identify, track, select) or whether DOD intends to have such a system. However, we should anticipate that potential opponents may seek to have such a system, which is one reason why the United States has reserved the right to respond in automated fashion to an opponent’s use of automated AI as a weapon.

A second question is if we are going to permit autonomous systems, who should be responsible (and held to account) for what the software does or does not do? We know from Stuxnet how a cyber-weapon might be employed, what it can accomplish, and how it might jump the rail, even when it is intended to operate and remain in an air-locked space. What then might a fully autonomous AI weapons system do and how, if at all, should the law seek to cabin that potential? That is the subject of chapter 23.

AI will also make it easier to test and develop weapons. As discussed in chapter 23, the law of armed conflict requires that new weapons as well as the means and methods of warfare be tested for compliance with the law of armed conflict prior to deployment. AI capacities can be used to model this activity, where actual testing is unfeasible, unreliable, or incomplete. Nuclear weapons, which are already tested through modeling, are an obvious example. But so are cyber weapons and swarms. Just as pilots train through simulation, AI can simulate variables that will impact weapon performance.

But wait there is more! If the focus has been on AWS, AI’s military impact will likely be as great on military intelligence, decision-making, and logistics.

### B. Intelligence

AI’s immediate promise as a national security tool is likely most evident in the area of intelligence. Here narrow AI’s capacity to outperform human capacity in pattern recognition, anomaly detection, and the speed with which it can do so is evident. Just think of the application of AI to intelligence concepts like “connecting the dots” and the “mosaic theory.” Or, imagine the computational capability to search the entirety of the Internet for threats in real time, along with a capacity to analyze those threats for links to purchase, travel, and phone number patterns. AI also provides the capacity to immediately analyze information, or a story, and to trace its origin in cyberspace to the head of the stream and thus help to distinguish between fact and fiction, or something between. This is as true in military context across domains as it is for national intelligence.

More specifically, AI software or enabled systems can, or will, manifest themselves in the following specific applications.

- Persistent surveillance
- Image Recognition, including facial recognition
- Voice Recognition
- Sorting
- Aggregation, a.k.a., Fusion
- Political prediction
- Translation
- Deviation and anomaly detection
- Cyber detection, attribution, and response

Now let’s consider how these capabilities might impact some of the intelligence disciplines.

**Analysis:** It is axiomatic in intelligence analysis that generally there is either too much information to analyze or too little. The latter might be illustrated by the challenge of determining intent within the leadership circle of a closed authoritarian regime, like North Korea’s, or the Soviet Politburo, or one with a small footprint, like a terrorist cell or the government of South Sudan. The former might be illustrated by
just about everything else, but for the sake of brevity, consider the complexity of Open Source Intelligence analysis in the context of the Internet, including the Dark Web.

One might think that the dilemma of too much information is a recent phenomenon, but the challenge is not new, the scale is. Sherman Kent, one of the architects of intelligence analysis as a discipline, wrote in the 1950s about the overwhelming amount of information potentially available to analyze and the necessity of knowing when and where to put the man in the informational loop. (That was in the 1950s!) Of course, the magnitude of the challenge is greater today. Pick your metaphor, analysts talk about noise to signal ratios or finding a needle in a haystack. Analysts used to measure the amount of data collected with reference to the number of Libraries of Congress. Today a common unit of measure is the petabyte, the equivalent of the holdings of seven Libraries of Congress. And, by the time this chapter is published, information may be routinely sorted into exabytes. Intelligence specialists have also spoken for years of the mosaic theory of intelligence and, after 9/11, we all know about the necessity of connecting the dots, the process of piecing together diffuse pieces of information to create a greater whole for the purpose of informing, warning, and predicting.

AI is an intelligence force multiplier. It is already better than humans at many of the analysis tasks, like sorting, categorizing, pattern recognition, anomaly detection, and certain types of link analysis. Properly constituted algorithms can search the Internet. They can detect anomalous trade patterns or travel patterns. If one were tracking sanctions enforcement and evasion, AI is an invaluable tool. It can find, aggregate, sort, and identify anomalous patterns in the transfer of goods based on bills of lading, bank transfers, shipment weights, routes and all the other data which lies beyond the capacity of human fingertips to collect and analyze in real time or near real time.  

**Fakes, Ruses & Covert Action:** AI’s natural language capability and pattern analysis processing that can be used to identify voices can also be used to mimic voices, alter images, and mock code, referred to sometimes as fakes or deep fakes. This has potential, perhaps already realized, to enable some of the traditional tools of covert action, in American parlance, or direct action, in Russian parlance, like the creation of disinformation, false flags, and propaganda activities. We can see from the February 2018 indictment of thirteen named Russian agents for interfering in the 2016 U.S. presidential election just how aggressive Russian use of cyber instruments as covert tools can be. The Russian government operated a twenty-four-hour-a-day bot farm, spreading false flag information not just about the presidential candidates, but seeking to suppress the vote in African American communities. The Russian efforts extended across social media platforms—Twitter, Instagram, Facebook. One Russian social media account had over 100,000 followers. AI can make cyber weapons more

---

Effective, by helping to find zero days and by enhancing the speed of exploitation and by disguising attributable characteristics of the attacker.

In military context, efforts to deceive the enemy are often called ruses. The Trojan Horse was a ruse. So were the subterfuges used by the Allies to keep the Germans guessing as to where the D-Day landings would occur, including metal chaff to mimic aircraft formations, mock aircraft and tanks, and an irate General Patton commanding fake armies from Scotland and broadcasting in the clear to do so. AI will allow military forces to engage in such ruses more effectively, perhaps dangerously so. AI might be used to disable an opponent’s air defense system, or perhaps take it over to be used against the opponent’s own aircraft. AI can be used to mimic the command voice and code of an adversary’s chain of command, and realistically so. And, AI can be used to mimic a nation’s leaders to sow confusion as to home front views and politics and undermine morale. All of which heightens the need for sound encryption and active counterintelligence.

**Counterintelligence:** In the area of counterintelligence, AI will have, and no doubt has had, two immediate impacts. First, it is a counterintelligence enabling tool. It can aggregate information and identify patterns in financial, physical, and digital behavior along with any anomalies in that behavior. Consider how quickly your credit card company knows when your credit card is used out of pattern, for example, when you travel overseas, or purchase gas on a long distance road trip. One need not be a specialist to appreciate how such analysis might help identify a Snowden or a Manning accessing information outside their normative responsibilities, or an Ames or a Hanson, spending money in new ways or beyond their apparent means.

But CI cuts both ways. In converse fashion if it might be easier to find the spy or asset, it is also harder to run or place the asset or clandestine officer to begin with without providing an electronic data trail. Likewise, if AI enables counterintelligence, it also enables the internal police to track citizens more effectively within the construct of an authoritarian state, counterintelligence of a different sort.

Second, if AI is a transformative military technology then it will, by definition, become an intelligence target and CI challenge commensurate with its importance. Because a majority of AI research and development is occurring in academic and corporate laboratories, these same laboratories will be intelligence targets in a way they have not been before. Federally Funded Research and Development Centers (FFRDCs), as well as Google and Facebook, are new targets requiring new efforts to spot and counter technical and human penetrations. Likewise, cyber data used for machine learning will take on added importance as an offensive target to inform AI applications. One need not be a machine learning specialist to appreciate that the more data an algorithm is trained with the more accurate and less bias the application may reflect. That makes new kinds of data as well as old kinds of data valuable in new ways to inform AI driven link analysis. Consider how the SF-86 data stolen from OPM by the Chinese could be used for machine learning or for that matter the ubiquitous effort to collect DNA to compose family trees and identify ancestry, most of it going to China for processing. We should not be surprised to find this same data making its way into deep learning-enabled machines for further analysis and later pattern matching.
One question for the military specialist is how to deploy AI-enabled applications down to the tactical level in a manner that appropriately mitigates the greater risk of counterintelligence penetration or loss such deployment creates.

**Targeting.** With military targeting, AI’s impact on intelligence and weapons comes together. Consider AI software that can truncate hours of surveillance video from a UAV feed into minutes, ensuring key facts and events are not missed. This is one purpose of the DOD Maven program, which led some Google employees to protest the company’s association with the project. Such a processing system can also eliminate the sorts of human mistakes that come with fatigue, repetition, and from what Paul Slovic calls the prominence effect, the cognitive tendency to focus on the mission and immediate objective with the unintended consequence of excluding other variables.

**C. Logistics**

Other military uses will derive from AI’s human level capacities, and in many cases better than human level intelligence (BHLMI). These include:

**Training & War-gaming.** AI can enable realistic simulated training for pilots and other military actors. It can also be used in war games and exercises. Consider, if an AI-enabled computer can play Go or chess and beat humans doing so, it can simulate a military opponent on a tactical or strategic level providing additional human training opportunities.

**Translation.** A self-evident capability that could potentially provide “translators” down to the lowest level of maneuver element. This offers significant advantage in counterterrorism and counterinsurgency contexts where the support, or at least neutrality, of local populations is essential and miscommunication disastrous.

**Logistics.** AI-enabled software and computational capacity can model the best methods to transport and deliver logistics, taking into account weather, fuel, need, and any other factors that might take an inordinate amount of time for humans to calculate. Consider what a Waze or GPS-activated navigational system would act like if it took into account all relevant factors in assessing the best route to deliver supplies from Point A to Point B. Even better, imagine planning D-Day or the 1990 Desert Shield deployment with the benefit of AI algorithms that can plan down to the second, optimum airlift and sealift schedules to address needs and contingencies, and instantly adjust these schedules every time the weather shifts, or an ally adjusts its end force commitments. The Defense Department has also described how AI will enable logistics delivery and not just planning. “Elevated levels of autonomy in unmanned systems will allow for leader-follower capabilities, where trailing semi-autonomous vehicles follow a designated vehicle in logistics convoy operations.”

---

D. Decision-Making

Each of these capacities come together to inform and potentially change how certain military decisions will, or could, be made. This may occur in generally positive ways. For example, if more intelligence, and more accurate intelligence, helps policymakers better predict threats and thus better deter threats before they materialize, then AI’s capacity to identify, fuse, and connect intelligence streams is a positive development, if it is used and used wisely. Likewise, any process that can more accurately and rapidly distinguish between real and fake information, and between noise and signals, should contribute to better decisions. In theory, intelligence also contributes to stability by reducing the risk of miscalculation and misperception, at least where the intelligence is accurate and understood. Of course, more intelligence also means more noise, and thus a necessity to adjust any decision-making process to account for this problem.

If used effectively, AI will also help decision-makers model potential policy outcomes, just as Watson rapidly modeled Gary Kasparov’s chess moves and countermoves. It will also allow Red Team and Blue Team challenges to policy proposals in real time of the sort that General Van Riper displayed in war games involving the Persian Gulf described in Malcolm Gladwell’s book *Blink*. But these are all potential benefits. They only become actual benefits if the policy process is changed or adjusted to effectively provide this input at the tactical and national level. This will require new expertise and new participants in the ordinary command processes. It will also demand that commanders know when to rely on AI to augment human capacity and decision, when to defer to AI systems outright, and when to rely on human intuition alone.

AI will also introduce new risks and exacerbate existing national security decision-making pathologies, such as those associated with the rapidity of decision, secrecy, and incomplete facts, which is the subject of the next section.

III. PROBLEMS AND CHALLENGES

The potential benefits of AI also bring potential problems and challenges. These potential challenges are tactical, like ensuring AI capabilities are used wisely as intelligence indicators, or, that operators interface correctly with AI-enabled systems. The 2017 MITRE Corporation JASON Report (p. 2), “Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD,” identified some of the risks of AI R&D and deployment as the “so-called ‘ilities,” which “are of particular importance to DoD applications.”

“Reliability
Maintainability
Accountability
Verifiability
Evolvability
In the so forth category, one might also include, for example, interoperability. These risks are largely tactical; however, they soon become strategic if tactical risks play out on a strategic geopolitical stage, in the Black Sea or South China Sea, for example. There are other strategic challenges as well, such as the consequences of an AI arms race and defining the doctrinal principles that should govern the use of AI for national security purposes.

This section introduces the reader to some of these risks. For purposes of structure and brevity, one might frame the risks and challenges using the following four rubrics: (1) National security decisional pathologies; (2) the “law” of unintended consequence and the man-in-the-loop dilemma; (3) foreign relations and internal stability; (4) arms race risks and imperatives.

A. Decision-Making Speed

National security decision-making is subject to a number of process pathologies. In this context, a pathology is defined as a factor or condition that undermines optimal national security decision-making. Good process, meaning timely, contextual, and meaningful process, along with calm leadership, is the antidote to the decisional pathologies. However, they are mitigating antidotes, not eliminating immunizations. Five prevalent pathologies are:

- Speed
- Secrecy
- Incomplete and/or lack of complete information
- Focus on the immediate
- The national security imperative.

Issues of cognitive bias also come into play, especially in how decision-makers analyze information, assess history, and apply doctrinal perspectives.

AI has the potential to exacerbate, or mitigate, each of these decisional pathologies. This section starts with the rapidity of national security decision-making, or speed, because speed is a signature attribute of most national security AI applications. It is also a factor that can aggravate most of the underlying AI risks, including those associated with the decision-making process. It is certain to do so, potentially pushing decision-makers to remove humans from in or on the loop of decision, as well as take R&D shortcuts to save time and keep up.

Rapid decisions are endemic to national security. The compulsion and necessity for speed comes from a number of factors. In the case of real world events, the need for speed is intuitive. If you are reacting or seeking to influence events, your timeline is dictated by those events and not by optimum considerations of process, factual development, and policy consideration. Moreover, opponents may seek moments of distraction and commitment to act presenting additional security challenges and
further minimizing the capacity and time to respond. The intelligence process seeks to prevent surprise and to provide early warning, and thus extend decisional timelines and mitigating opportunities. But intelligence is an imperfect instrument for reasons of scope, capability, and the simple difficulty of predicting or seeing actions that are designed to be hidden.

Technology has also changed the decision-making timeline, and not just in the handling of public communications. The most dramatic manifestation of this trend is in the realm of cyber operations. Cyber tools used for hacking, crime, espionage, and covert action are instantaneous in their use and often in their effect. That means that technology-based defenses must be instantaneous in their response as well, if they are to be effective. That means that in cyberspace, decision-makers are always on the clock and this clock runs on milliseconds. Thus, it is no surprise that one of the defining terms in cyber operations—zero days—is not only time-based, but loudly declares to all who know, “You are on the clock.” There have been zero days since this vulnerability was discovered (and is now being exploited); how many more days will there be before you find it and fix it?

AI will, or can, depending on how it is applied, both mitigate the impact of speed on decision-making as well as exacerbate it. It can do so in profound ways. A decisional process that is not ready for these impacts will underutilize the advantages of AI as well as rue the advent of AI’s disadvantages. What are they?

The combination of algorithmic intelligence and computational speed can help decision-makers identify, fuse, and connect relevant information faster and more effectively than humans. This is the core strength of narrow AI at this time—pattern identification, link-analysis, and speed. Therefore, as already noted, AI can also help identify the sort of anomalies that serve as intelligence flags and warnings. This creates more time for decision-makers to process information, frame options, and make decisions. Consider that many lawyers and firms are already using AI link analysis and pattern recognition to perform discovery searches and sorting, saving time and money, and advocates argue, improving accuracy.

At the same time, there are added risks generated by speed. Because AI can move so quickly, if deep learning is effective, in some situations decision-makers may have less time or no time, (or perceive that they have less time or no time) to respond. This can drive policymakers to rapid decisions or to defer to automatic responses, which may or may not be optimally tailored to the actual scenario at hand. This is, of course, already an existing reality in the realm of cybersecurity and cyber operations. In cyberspace, there is risk in waiting to respond to an attack while facts are gathered, attribution confirmed, and options identified. This paradigm may drive automatic responses to defensive options and away from offensive-defensive or offensive responses that may more effectively stop attacks and serve to deter future attacks.

1. Other Impacts

However, there are a number of factors that make the speed introduced by AI-influenced decision-making different from the sort of decision-making that must already occur in relation to cyber events.
First, for the reasons stated in the previous section, AI will influence the intelligence function, which function shapes options and informs national security decisions across the board.

Second, AI will also be linked to the deployment and potential use of kinetic weapons. Depending on how this is done, AI could drive a policy of immediate response with regard to kinetic weapons, just as cyber weapons have driven a policy of immediate response in cyberspace, unless of course, decision-makers build in fail safes and human circuit-breaking. But whereas the threat of a nuclear first-strike usually is estimated to leave decision-makers with a 15–20 minute window (less in the case of a submarine launch) to assess whether there is an attack underway and if so how to defend and respond; with AI a first-strike may be instantaneous. Thus, there is need for new deterrent strategies and doctrine to account for the speed with which a first-strike, in theory, could occur with AI-enabled cyber weapons.

Third, AI may convey a sense of mathematical confidence, or lack of confidence, to decision-makers that may prove false in context. The making of foreign policy and national security decision-making is an art, not a science. National security policy cannot be reduced to a mathematical formula. Moreover, even if it could, one's opponent could trick or spoof the formula. Rather, national security policy-making entails fact, but also intuition, historical knowledge, values, and judgment about reactions and counter-reactions. If national security were math, we would be more afraid of driving than terrorism. Statistically driving is more dangerous. That is why the risk of terrorism and the cost of preventing terrorism cannot be reduced to a law and economics model based on efficiency. People confronting fear will give prominence to that fear, whether it is rational and math-based or not. And it is the effect of that fear that policy seeks to control, not just the predicate acts of terrorism.

Henry Kissinger said that most issues that come to the President are 51–49 issues; in other words, close calls and judgments. Otherwise, they likely would not come to the President. But most decisions do not lend themselves to formulaic response. Deep machine learning, at present, is computational, and thus has a tendency to classify facts or options that present policy choices into mathematical probabilities, either expressed in terms of a conclusion (there is a 76 percent chance this is a tumor), or expressed in terms of an accuracy of instrument (AI-enabled radiology analysis is correct 76 percent of the time). Some decision-makers covet these judgments because it makes hard decisions easier; who can rightfully object if there is a 76 percent chance of a policy action working. For these reasons, intelligence analysts—and DNIs—are constantly pushed to put a percentage on their judgments. These same actors resist doing so, resorting instead to judgmental phrases like "we assess" and weight judgments with qualifiers like "likely" and "unlikely." Some judgments and decisions just do not lend themselves to precise predictions.

There can be comfort in numbers. But there is also decisional risk on both a tactical and strategic level. Can a decision to target an individual combatant be reduced to a formula? And what if intuitive situational awareness suggests a less certain reality? Likewise, some of the great judgments in history would not appear to lend themselves to computational, formulaic, or percentage responses.
How would the Dunkirk evacuation operation have looked if the logistics were first run through AI applications or the overall possibility of success or failure classified by a percentage? Would it have been possible to compute the risk to morale of not trying? Or, what percentage of men on the beach were needed to prevent a subsequent invasion of the UK?

AI would certainly have helped predict the weather on June 6, and for the coming weeks. But would it have made the decision to go, or not go, any less instinctual?

And, how would AI have influenced President H.W. Bush’s decision not to push on to Baghdad at the close of the First Gulf War? No doubt the computations would have indicated a favorable prognosis for success, making it all the harder to say no.

There also seems like there is something qualitatively and morally different between a decision to strike a target with a Hellfire missile based on an AI probability that is 76 percent sure the person is a terrorist versus the informed judgment of the DNI that the person is a terrorist posing an imminent threat to the United States.

Fourth, with machine-driven predictions, such as those generated by neural networks, decision-makers may not fully understand the basis of their own policy choices and decisions. This will make it harder for decision-makers to explain their choices, or it may lead decision-makers to forsake the use of a valuable security tool to augment the human decision-making function.

Daniel Kahneman received a Nobel Prize for his work identifying the differences in brain function between what he described as fast and slow thinking. In brief summary, fast thinking is automatic, frequent, emotional, and unconscious. Slow thinking is logical, calculating, and conscious. National security decision-making requires both. AI at present, seems especially suited to augment national security decision, not supplant it. A timely, contextual, and meaningful decision process, should and must incorporate AI functions, but also know when and how to use AI output wisely as well as have the discipline and courage to resist over-reliance, or under-reliance, on its output.

B. The Law of Unintended Consequences and the Challenges of Machine-Human Interface

Technology does not always work as intended. Here we are not talking about research and development miscues and mistakes, but deployed systems. Some examples are funny, or perhaps only apocryphal; Icarus’s wings did not work as intended. Most examples are neither funny nor apocryphal. The Mark-19 Torpedo, the principal attack torpedo in the Navy’s pre-World War II inventory had a tendency not to explode upon contact and in some cases returning (it is thought) to strike the submarine from which it was launched. America’s first two satellite launches following Sputnik exploded on launch. The Thresher and Scorpion and Apollo 13 are testimony to technology that worked well and then failed to work as intended. Let’s add Columbia and Challenger
to this list. There are also examples of technological emergencies that were supposed to occur, but did not. Remember the Y2K crisis, which some predicted and the government feared might lead to a cyber-meltdown? There is hubris in thinking AI will work precisely as intended. Remember, many current narrow AI applications depend on classifying data and making probability assessments about that data. Chatham House scholar M.L. Cummings describes this process with respect to driverless cars.

Indeed, this form of estimating or guessing what other drivers will do is a key component of how humans drive, but humans do this with little cognitive effort. It takes a computer significant computation power to keep track of all these variables while also trying to maintain and update its current world model. Given this immense problem of computation, in order to maintain safe execution times for action a driverless car will make best guesses based on probabilistic distributions. In effect, therefore, the car is guessing which path or actions is best, given some sort of confidence interval. The best operating conditions for autonomous systems are those that promote high-fidelity world model with low uncertainty....

But what if the assessment is wrong? With a car, the consequences are presumably finite and limited. Where AI is enabling a weapons system or warning system the consequences could be catastrophic.

The risk with strong AI and AWS, commentators like Nick Bostrum believe, is that you may only have one shot to get the technology right, if indeed the AI-enabled technology, capacity, or weapon, is introduced to the Internet. It is one thing if it fails to successfully perform its task. But what if it does not shut down as intended? Or, continues to pursue its goals, through other machines when commanded to stop, or disconnected from the Internet? Experts express some skepticism about how realistic this scenario really is.

But how farfetched is it to ask whether AI will work as intended? What if Stuxnet, for example, was introduced on to the Internet and not into a confined system, and what if it could not only attack a SIEMENs designed SCODA, but was designed to identify firewall defenses, and rewrite its code to find an entrance? One potential lesson from the Cold War arms races is that arms races place pressures on states and actors to produce and match, to deploy before ready, tested, and full proof. The “Get it Right Once” risk might occur in a different kind of setting, where AI is a critical component in a tactical or strategic warning system, for example, a carrier born system designed to identify and respond to a swarm attack or a cruise missile attack; or a satellite system designed to identify a missile launch.

One way to address unintended consequences is to ensure there is a person in the decision loop to act as a circuit breaker. This seems particularly imperative where weapons are concerned. Here, one of the questions is just where to put the man in, or in the current vernacular, on or out of the loop. However, part of the allure and the advantage of AI-enabled systems is their speed and ability to respond in instantaneous manner, whether responding to swarms, cyberattacks, missile launches, or
stock trades. The tension is between taking full advantage of the AI capacity, and providing a human air gap in the system—on or in the loop—that will slow it down and introduce both human judgment (and human frailty) into the decisional-loop. There are military and arms races pressures pushing, perhaps inexorably, toward out-of-the-loop concepts.

In this setting, the risks may come in threes.

1. AI appears to work as intended, but the human in the loop does not understand the results.

2. The human in the loop does not have time to process or act on the results.

Or,

3. The AI-generated results are not accurate. We know this latter risk from the Cold War, and on both sides of the equation.

Sydney Freedberg and Matt Johnson have illustrated some of these risks with reference to Air France flight 447, which crashed in the Southern Atlantic in transit from Brazil to France in 2009. Apparently, the pilots were unable to transition from autopilot to manual flight in the midst of an in-flight emergency allowing them only seconds to respond. The aircraft stalled and crashed into the ocean. The authors cite as well a 2003 friendly fire incident in Iraq involving a Patriot battery. The technology worked as intended, but when the technology passed control back to the humans in the loop to make a fire or no-fire decision, the operators were not sure what they were seeing and made an erroneous choice to fire.14 Likewise, the Vincennes incident, involving the shooting down of a civilian Iranian airliner by an Aegis cruiser in 1988 has been studied as an illustration of where the technology worked—the data was correct regarding the speed, direction, and climb of the aircraft, but in the moment, human actors misread the data, perhaps because of the pressures of the moment and perhaps because the command was predisposed to perceive a threat.

Now connect these risks to existential weapons. In 1979, in an incident documented on the National Security Archives website and in Robert Gates’s 1996 book From the Shadows, President Carter’s National Security Advisor Zbig Brzezinski was awakened at three in the morning by [military assistant William] Odom, who told him that some 250 Soviet missiles had been launched against the United States. Brzezinski knew that the President’s decision time to order retaliation was from three to seven minutes. Thus he told Odom he would stand by for a further call to confirm Soviet launch and the intended targets before calling the President. Brzezinski was convinced we had to hit back and told Odom to confirm that the Strategic Air Command was launching its

planes. When Odom called back, he reported that 2,200 missiles had been launched. It was an all-out attack. One minute before Brzezinski intended to call the President, Odom called a third time to say that other warning system were not reporting Soviet launches. Sitting alone in the middle of the night Brzezinski had not awakened his wife, reckoning that everyone would be dead in half an hour. It had been a false alarm. Someone had mistakenly put military exercise tapes into the computer system.

History, in a way, would repeat itself four years later. This time, Soviet Lieutenant Colonel Stanislav Petrov was serving as a watch officer at a Soviet early warning radar and command center. Alarms went off indicating that an American first-strike had been launched at the Soviet Union. Petrov was skeptical. The pattern on the radar did not look like what he anticipated a first-strike would look like. There were too few missiles. As the senior officer of the watch, it was his duty to call the Kremlin to trigger the Politburo’s response and perhaps initiate a nuclear exchange. He stalled. As recounted in his 2017 obituary in *The New York Times* quoting a BBC Russian service interview thirty years after the incident.

I had all the data [to suggest there was an ongoing missile attack]. If I had sent my report up the chain of command, nobody would have said a word against it. There was no rule about how long we were allowed to think before we reported a strike. But we knew that every second of procrastination took away valuable time; that the Soviet Union’s military and political leadership needed to be informed without delay. All I had to do was to reach for the phone; to raise the direct line to our top commanders—but I couldn’t move. I felt like I was sitting on a hot frying pan.

The warning system squawked alert; data pointed in one direction, but with an anomalous pattern. Petrov’s military and bureaucratic training should have driven any doubts and decisions up the chain of command. But Petrov’s intuition suggested otherwise. He took the risk and was right. But not every lieutenant colonel is a Petrov and not every national security advisor a Brzezinski.

With AI, there is a likelihood that there will be less time to adjust, if the system is intended to augment human decision, unless of course, the system is autonomous and automatic. Recall that what AI is less good at than humans is situational awareness and judgment, which depend on context, experience, and intuition. It is also more likely that an AI-enabled system will instill greater confidence in its operators than Cold War era technology, and likely with good reason. But in the abstract, would you be more likely to trust a Soviet era early warning system or IBM’s Watson. Moreover, remember what narrow AI is best at, and better than humans at, pattern recognition and identification. That is what early warning detection is all about. However, another question might be, would you be willing to trust Watson—to bet your life on Watson—would you bet the globe—without first having a trained, calm, and rational officer first assess the results?
Recall that Brzezinski not only knew enough to seek initial verification of the initial attack call, he appears to have calmly waited as the minutes ticked by for not one, but two verification calls. In other words, he was in the loop, and waited in the loop as the clock ticked down. Note as well that according to the National Security Archives the problem was not the mistaken use of an exercise tape, but the loading of software into NORAD's computers. “The information on the display simultaneously appeared on the screens at SAC Headquarters and the National Military Command Center (NMCC)...” thus erroneously confirming the attack.

Arms race pressures, to deploy better and faster systems and to maximize the advantages of AI, may however remove the man from the loop. In something of a reverse no-first-use-pledge, the United States has stated that it will not deploy weapons systems without a man in the loop, unless an opponent does so first, thus creating a tactical or strategic advantage. Other states have not made similar public pledges.

The risk of unintended consequences is compounded by three counterintelligence risks. As the DOD Roadmap report notes, “This problem is especially apparent in unmanned systems, which by their very nature have an elevated reliance on information systems to function safely, effectively, and consistently.” [Roadmap 24] The first risk is from supply chain contamination or sabotage. The AI-enabled system will not work as intended because an opponent has altered the circuitry or code involved, by introducing faulty hardware or software into a weak link in the supply chain. The second risk comes in the form of the intentional use of AI as a weapon, perhaps through supply chain weakness or through AI first-mover advantage to cause an opponent’s system to fail. Finally, AI presents new capabilities to spoof an opponent by disguising an attack, camouflaging attribution, or engaging in false flag operations.

Finally, a technology arms race with respect to AI is likely to prompt a parallel espionage race. The faster the race occurs, the faster the effort to collect information on that race and to curtail an opponent’s advantage through theft and espionage. Knowledge of an opponent’s capabilities and intentions can be stabilizing, as in the case where actual knowledge debunks perceptions of a bomber or missile gap and thus spurs unnecessary and additional arms expenditure. However, it can also be destabilizing where it leads to increased, if not rampant, efforts to penetrate and steal an opponent’s knowledge and capabilities, which in the case of AI may also lead to uncertainty over the integrity of any ensuing AI function. In this race, one of military and economic espionage, the United States may find itself at an asymmetric disadvantage depending on how it approaches the subject of economic espionage in law and policy.

C. Foreign Relations Impact and Instability

AI will have foreign relations impact in at least eight ways. First, it will impact global stability in known and unknown ways, just as climate change will do so. As quoted in the Belfer study, former Secretary of the Treasury and Harvard economist Larry Summers, for example, predicts that on a global basis we “may have a third of men between ages of 25–54 not working by the end of this half century.” This would represent a higher unemployment rate and number than during the Depression,
which could affect political as well as economic stability and lead to mass migration, geopolitical instability, and military conflict.\textsuperscript{15}

Second, a global AI economy will also potentially reorder our understanding of north-south divides, third-world, second-world, first-world orders, perhaps by exacerbating those divides. And, of course, AI capacity may redefine the nature and number of superpowers, a phrase first coined to capture the advent of the nuclear weapon.

Third, AI could alter and shuffle the relative power of small, but technologically sophisticated states in terms of economic, political, and military power. Call this the Singapore effect. To the extent AI comes to influence, or perhaps, dominate the nature of military power, it may give smaller less populous states the military capacity to fight above their weight or do so in an asymmetric manner.

Fourth, to the extent military power, intelligence capacity, and economic stability is hinged to AI capacity, it will increase the importance of supply chain reliability. The axis of national security issues that revolves around nonproliferation import and export supply chains—including like-minded trade regimes—may increase in importance. AI-enabled technology will only be as good or reliable as the individual components that comprise AI-enabled systems like transistors, circuitry, and software. A failure to advance AI capacities could have devastating national security impact. So too could reliance on an AI-enabled capacity that has been penetrated by an opponent.

Fifth, AI presents asymmetric opportunities for non-state actors in the same way that it creates those opportunities for state actors. Unlike nuclear weapons, which present a triad of obstacles to their acquisition, AI is potentially accessible to almost every actor at a relatively inexpensive price, depending on how it is defined and what it is used for. Consider, for example, that the autonomous commercial vehicle or remotely piloted aerial delivery system can also be used as a mobile IED or bomb. Consider as well that the Stuxnet code jumped the rails, was publicly identified by a private actor, and could potentially be reused.

Sixth, AI may help authoritarian regimes better track and control their populations and thus facilitate the retention of power. AI algorithms are the censor’s tool on the Internet. Gregory Allen, co-author of the Belfer/IARPA study on “Artificial Intelligence and National Security” describes how facial recognition can be used as well.

Snapchat uses AI-enabled facial recognition technology to allow teenagers to send each other funny pictures. China uses the same technology in support of domestic surveillance. Jaywalk across a street in Shenzhen, and you’re liable to have your face and name displayed on a screen nearby, along with a police reminder that “jaywalkers” will be captured using facial recognition technology.\textsuperscript{16}

\textsuperscript{15} Belfer/IARPA, p. 2 and p. 37.

Hold your cellphone to your ear while driving on certain segments of the M-4 highway in the UK and a camera will automatically take a picture of the driver and send a notice of license suspension to the driver’s address. Advantage goes to the authoritarian regime and to law enforcement.

Finally, the advent of AI will likely lead to an arms race between powers. It already has. Commentators note that AI-enabled systems are particularly suited for the vast spaces presented in the Pacific region and in so-called areas of denial, like the Western Pacific, read South China Sea. The national security imperative dictates this result. Some might wish this genie back in the bottle, but it is too late. So long as states perceive national security advantage from AI and as importantly existential disadvantage from the opponents AI capacity they will feel compelled to keep pace and respond in kind. The questions for policymakers, lawyers, and ethicists include:

- What lessons can be learned from the Cold War arms race?
- What principles from arms control and LOAC might or should apply to AI?
- What strategic and tactical doctrine should apply to AI enabled weapons, weapons systems, and warning components?
- What, if any, opportunity costs pertain to an AI arms race?

### D. A Technology-Driven Arms Race

Experts and commentators fear that the allure of autonomous weapons systems will lead to an arms race. Many leading AI researchers, for example, signed an open letter expressing concern about such a race.

Many arguments have been made for and against autonomous weapons, for example that replacing human soldiers by machines is good by reducing casualties for the owner but bad by thereby lowering the threshold for going to battle. The key question for humanity today is whether to start a global AI arms race or to prevent it from starting. If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow.17

However, an AI arms race is not only inevitable, it is already here.

As the Belfer Study indicates, and others have observed, AI is a potentially transformative military technology on par with nuclear weapons and the aircraft. It is also a transformative national security technology, which offers not just military advantage, but economic, intelligence, and decisional benefits. That means AI offers potentially decisive advantage to any breakout power in the field. That means this is a technology race and not just an arms race. For the United States, it presents the classic

---

17. Open Letter.
opportunity to offset, i.e., compensate for China’s advantages in geographic proximity to disputes arising in the Pacific region and numeric advantages in manpower and certain conventional weapons. It also offers the United States the allure of a panacea or technological fix to the challenge of maintaining a large standing military in a cost-conscious democracy.

But offset works both ways. For China, AI-enabled weapons, like swarms, hold the prospect of offsetting America’s naval advantages, especially in aircraft carriers, submarines, and ships, while AI-enabled cyber weapons may neutralize technological advantages the United States may hold in cyberspace. It also offers China an equal seat at the national security table in a way it was never equal to the United States or Soviet Union in a Cold War dominated by nuclear weapons and deterrence.

For Russia, AI-enabled weapons and direct measures offer a shot at renewed great power status, and the capacity to compete beyond its economic and military means. Russia’s willingness to use cyber tools to undermine its opponents, as evidenced by its interference in the 2016 U.S. election makes Russia’s interest in AI problematic. Given its potential to transform a nation’s physical and economic security, no power great or regional can afford to fall too far behind.

This arms race is also a race between systems and the relative security advantages of each. China has the advantage of centralized control and purpose. It also has an enormous cache of data. In 2017 China announced a $150 billion state driven program for AI development with a goal of becoming the leading AI economic and security power by 2030. In this system, the government can channel AI applications to national security applications without restriction. It can regulate and control data at the national level for security purposes.

The United States has the advantage of creative dispersion fueled by financial incentive and relative regulatory freedom. At present, six of the seven leading AI companies are located in the United States. But it should not be lost on American observers that Baidu, Tencent, and other companies are active partners in China’s efforts. Moreover, most commentators note that with AI, China has had its Sputnik AI moment. This occurred when AlphaGo beat China’s best player in 2016. China noticed. The United States has not had its Sputnik moment in AI. At least one former senior government official has stated, that if this were the Cold War we would be losing. Others offer a less bleak picture. Russia has the advantage of mission focus as well, perhaps, of low expectations. It also has the flexibility and freedom of action that comes from having less stake in the stability and viability of the international economic system and norms. Certain AI applications, offer great promise for a government willing to interfere through cyber means in the democratic and economic institutions of other states like Estonia, Georgia, United States.

There have been other arms races in history. Indeed, policymakers and decision-makers will search for metaphor to define and address the questions presented by an AI race. The most obvious metaphor is the nuclear arms race during the Cold War.

There are strong similarities, including to the time before nuclear weapons, when scientists and governments raced to harness the power of the atom, not quite sure how, and not quite sure when and to what end and ultimate result. AI, like nuclear weapons in the 1950s has the potential to transform military doctrine, spending, and policy. And, until that doctrine is set, understood, and stable, the world may be less stable.

However, there are many differences between AI and nuclear weapons, for one, AI is not a weapon, it is a range of capacities that can be used to enable weapons, robots, and autonomous vehicles and for many other purposes. Perhaps it is more like atomic energy, which can have both peaceful and military purposes. But two things that do seem more alike than different, is the potential for AI to transform military strategy like nuclear weapons did before, and the absence of a framework, let alone an agreed framework, to address the legal, moral, and ethical issues presented. Because security entails both physical protection and the preservation of our values, we should do so now. That is what it means to both uphold and defend the Constitution.

While nuclear weapons provides the most apt metaphor. Arms control more generally, and the law of armed conflict, offer additional lessons from which to assess AI. The threshold question now is which lessons are most apt.

IV. CONCLUSION

AI will transform intelligence and military systems and capabilities. Thus, it will also transform the national security decision-maker's toolbox. A greater number of automated and unmanned weapons selections may change the policy calculus in context for employing military force, just as remotely piloted vehicles have done so. The calculus may change, because force may be used with less risk of U.S. military casualties as well as less risk of civilian collateral casualties. Whether this is a net positive development is contested by commentators who see in AI-enabled warfare a risk that without "cost," at least cost to the initiator of conflict, policymakers may be more likely to resort to force. The same argument has been made with respect to UAVs. Of course, much but not all of the international law and policy is reciprocal and if that is the case, it may also be the case that AI-enabled weapons may increase the frequency in which such weapons are used against the United States and not just by the United States. These are not remarkable insights. They dominate the literature on AWS, along with debate over whether and when a human should or must be engaged in the lethal use of force.

These factors, including the focus on the immediate, assures that there will be an AI arms race. So long as AI in some form and in some manner holds out the prospect of military advantage, including existential military advantage, national security actors will strive to keep pace. They cannot risk doing otherwise. So long as AI holds out the prospect of first-mover advantage, or in nuclear parlance first-strike potential, it would be malfeasance for national security actors not to push forward. These same factors may also drive decision-makers to false optimism and can-do attitudes that
we will find a fix by the time we get there. If there is a fork in the road between friendly and unfriendly AI, we will choose friendly at that time.

The antidote to some of these risks is good process. Good process considers long term effects and not just immediate needs. One way to do this is to include a fuller set of actors at the decision table. Good process involves legislating to guide and not just to respond. Good process provides clear right and left boundaries of conduct. Good process also includes accountability and responsibility forcing decision-makers to make conscious accountable decisions, preferably involving more than default decisions.

A second antidote is to create and invest in a different sort of arms race. This race is intended to provide an ethical, legal, and doctrinal framework before any of the critical actors make irrevocable decisions. We can hope for the best, but we should not take it for granted. We should create it. Fred Kaplan describes in *Wizards of Armageddon* the creation of American nuclear doctrine following World War II. There was no doctrine to start for a weapon unlike any other. Hiroshima and Nagasaki notwithstanding, in the post-war years there was serious consideration given to the first use of nuclear weapons. Many strategists viewed atomic bombs, and even the thermonuclear weapons that followed, as just another generation of weapon offering extraordinary firepower. Civilian and military strategists debated the merits of using such weapons against cities and then as a deterrent or response to Soviet conventional incursions against the West. Now is the time to create such a doctrine as well as a legal regime to regulate military uses of AI.