**NEG**

### Inherency Resps

**The DOD has increased funding for TEVV and adopted standards for responsible AI**

**Atherton 2022 – Military Technology Journalist** [Kelsey, 5/6/22, “Understanding the errors introduced by military AI applications”, <https://www.brookings.edu/techstream/understanding-the-errors-introduced-by-military-ai-applications/>, 6/18/22, LND]

The Pentagon has taken some steps to address these risks. In February 2020, the Department of Defense released a set of [principles AI ethics](https://www.defense.gov/News/News-Stories/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/) drafted [by the Defense Innovation Board](https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF). One of these principles is “traceability,” emphasizing that relevant personnel will “possess an appropriate understanding of the technology,” including transparent and auditable data methodology. To foster that understanding and ensure that nondeterministic systems can be audited, the Pentagon is investing in testing, evaluation, validation, and verification methods for AI. The development of testing and explainability tools for military AI applications represents [one of the key challenges](https://www.brookings.edu/techstream/the-testing-and-explainability-challenge-facing-human-machine-teaming/) for the technology, and making the necessary investments to develop these tools will be key to responsibly deploying AI tools on the battlefield. This work is ongoing at the Joint Artificial Intelligence Center, which in February [awarded](https://www.c4isrnet.com/artificial-intelligence/2022/02/10/pentagons-ai-center-awards-contracts-to-79-companies-in-new-test-and-evaluation-agreement/) contracts to 79 vendors worth up to $15 million a piece to develop testing and evaluation technology.

At this relatively early stage of deploying AI in military applications, it’s important that researchers and policymakers develop what Holland Michel [describes](https://unidir.org/known-unknowns) as a “a finer-grain scheme for differentiating between different types of failure. By developing criteria to distinguish known unknown issues from unknown unknown issues,” policymakers can gain a more clear understanding of how AI systems are failing, which could “aid efforts to quantify risk in operations and assign due responsibility for unintended harm arising from data issues.” Another policy approach would be incorporating red teaming and adversarial assessment into the evaluation of AI products, as it would allow engineers of military AI to anticipate and plan for future failures in combat based on hostile action.

The additional challenges AI brings will come not in the existence of error, but in the nature of the error and the limits of explainability of the error. Thoughtful policymaking can anticipate this, and when in doubt, design systems that put machines in harm’s way before risking the lives of civilians or servicemembers.

**The JAIC increased support for TEVV**

**Konkel, 2020 - Executive Editor, Nextgov** [Frank, April 16 NextGov “ Pentagon Needs Tools to Test the Limits of Its Artificial Intelligence Projects” <https://www.nextgov.com/emerging-tech/2020/04/pentagon-needs-tools-test-limits-its-artificial-intelligence-projects/164687/> Acc 6/7/22 TA]

The Pentagon is shopping around for ideas from industry regarding how it might better test and evaluate future artificial intelligence products to ensure they are “safe and effective.” In a request for information this week, the Pentagon’s Joint Artificial Intelligence Center, or JAIC, seeks input on cutting-edge testing and evaluation capabilities to support the “full spectrum” of the Defense Department’s emerging AI technologies, including machine learning, deep learning and neural networks. According to the solicitation, the Pentagon wants to augment the JAIC’s Test and Evaluation office, which develops standards and conducts algorithm testing, system testing and operational testing on the military’s many AI initiatives. The Pentagon stood up the JAIC in 2018 to centralize coordination and accelerate the adoption of AI and has been building out its ranks in recent months, hiring an official to implement its new AI ethical principles for warfare. “The JAIC is requesting testing tools and expertise in planning, data management, and analysis of inputs and outputs associated with those tools. The introduction of AI-enabled systems is bringing changes to the process, metrics, data, and skills necessary to produce the level of testing the military needs, and that is why the JAIC is requesting information,” Dr. Jane Pinelis, Chief, Test, Evaluation and Assessment at the JAIC, said in a statement. “Testing and Evaluation provides knowledge of system capabilities and limitations to the acquisition community and to the warfighter. The JAIC's T&E team will make rigorous and objective assessments of systems under operational conditions and against realistic threats, so that our warfighters ultimately trust the systems they are operating and that the risks associated with operating these systems are well-known to military acquisition decision-makers."

### AT No Arms Race

**There Absolutely is an arms race - Russia and China are developing fully autonomous AI weapons now.**

**Sosanya, 2022 - AI researcher and a policy analyst at the Day One Project** [Andrew, Jan 3, Peace Review A Journal of Social Justice “Autonomous Weapons Are Here to Stay” <https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/10402659.2021.1998856> TM]

Russia recognized the inevitability of strategic competition with regard to autonomous warfare and has openly taken concrete steps to outpace its rivals. Chief of the General Staff of the Russian Armed Forces and Russia’s Deputy Defense Minister Valery Gerasimov announced that “robots will be one of the main features of future wars…[Russia] is seeking to completely automate the battlefield.” Russia’s Advanced Research Foundation (FPI) has publicly demonstrated one of its weapons in development: The Marker, an unmanned ground vehicle (UGV) equipped with anti-tank missiles and a machine gun, armed with its own swarm of UAV drones that can kamikaze into their targets. When asked about lethal decision-making, FPI Research Director Vitaly Davydov commented that everything will be under the control of a commander but “several options for the Marker are being planned out,” citing that he plans for the Marker to have fully autonomous capabilities. Autonomous weapons form an integral part in the Chinese military’s modernization plans. The People’s Liberation Army of China (PLA) has placed a scrutinizing eye on unmanned systems and autonomous capabilities to prepare for the warfare of tomorrow. China’s National Innovation Institute of Defense Technology (NIIT) has established two defense research centers that specifically focus on militarized AI, the Unmanned Systems Research Center (USRC) and the Artificial Intelligence Research Center (AIRC). China has developed and exported the Z3 Blowfish, an autonomous helicopter drone that can be outfitted with a machine gun or other projectile launchers, which is capable of performing coordinated swarm attacks in collaboration with other drones.

**Winning the AI arms race is key to maintaining US hegemony because AI gives our military Game Changing advantages, and China and Russia are right behind us.**

**Nurkin and Konaev, 2022 - senior fellows at the Center for Strategy and Security at the Atlantic Council** [Tate and Margarita, 5-25-2022, “Eye to eye in AI: Developing artificial intelligence for national security and defense” Atlantic Council <https://www.atlanticcouncil.org/in-depth-research-reports/report/eye-to-eye-in-ai/> ARD]

Over the past several years, militaries around the world have increased interest and investment in the development of artificial intelligence (AI) to support a diverse set of defense and national security goals. However, general comprehension of what AI is, how it factors into the strategic competition between the United States and China, and how to optimize the defense-industrial base for this new era of deployed military AI is still lacking. It is now well past time to see eye to eye in AI, to establish a shared understanding of modern AI between the policy community and the technical community, and to align perspectives and priorities between the Department of Defense (DoD) and its industry partners. Accordingly, this paper addresses the following core questions. What is AI and Why Should National Security Policymakers Care? AI-enabled capabilities hold the potential to deliver game-changing advantages for US national security and defense, including ● greatly accelerated and improved decision-making; ● enhanced military readiness and operational competence; ●heightened human cognitive and physical performance; ● new methods of design, manufacture, and sustainment of military systems; ● novel capabilities that can upset delicate military balances; and ● the ability to create and detect strategic cyberattacks, disinformation campaigns, and influence operations. Recognition of the indispensable nature of AI as a horizontal enabler of the critical capabilities necessary to deter and win the future fight has gained traction within the DoD, which has made notable investments in AI over the past five years. But, policymakers beyond the Pentagon—as well as the general public and the firms that are developing AI technologies—require a better understanding of the capabilities and limitations of today’s AI, and a clear sense of both the positive and the potentially destabilizing implications of AI for national security. Why Is AI Essential to Strategic Competition? The Pentagon’s interest in AI must also be seen through the lens of intensifying strategic competition with China—and, to a lesser extent, Russia—with a growing comprehension that falling behind on AI and related emerging technologies could compromise the strategic, technological, and operational advantages retained by the US military since the end of the Cold War. Some defense leaders even argue that the United States has already lost the military-technological competition to China.1

#### Winning the arms race is key to peace - AI enhances deterrence by improving military operations – it lifts the fog of war, accelerates decision making and saves soldiers’ lives

Nurkin and Konaev, 2022 - senior fellows at the Center for Strategy and Security at the Atlantic Council [Tate and Margarita, 5-25-2022, “Eye to eye in AI: Developing artificial intelligence for national security and defense” Atlantic Council <https://www.atlanticcouncil.org/in-depth-research-reports/report/eye-to-eye-in-ai/> ARD]

AI embodies a significant opportunity for defense policymakers. The ability of AI to process and fuse information, and to distill data into insights that augment decision-making, can lift the “fog of war” in a chaotic, contested environment in which speed is king. AI can also unlock the possibility of new types of attritable and single-use uncrewed systems that can enhance deterrence.2 It can help safeguard the lives of US service members, for example, by powering the navigation software that guides autonomous resupply trucks in conflict zones.3 While humans remain in charge of making the final decision on targeting, AI algorithms are increasingly playing a role in helping intelligence professionals identify and track malicious actors, with the aim of “shortening the kill chain and accelerating the speed of decision-making.”4 AI development and integration are also imperative due to the broader geostrategic context in which the United States operates—particularly the strategic competition with China.5 The People’s Liberation Army (PLA) budget for AI seems to match that of the US military, and the PLA is developing AI technology for a similarly broad set of applications and capabilities, including training and simulation, swarming autonomous systems, and information operations—among many others—all of which could abrogate the US military-technological advantage.6 As US Secretary of Defense Lloyd Austin noted in July 2021, “China’s leaders have made clear they intend to be globally dominant in AI by the year 2030. Beijing already talks about using AI for a range of missions, from surveillance to cyberattacks to autonomous weapons.”7 The United States cannot afford to fall behind China or other competitors.

### --Extend – China is Racing

**China is currently outpacing the US in AI competition – they have invested in jointness capabilities and autonomous drone swarms**

**Nurkin and Konaev, 2022 - senior fellows at the Center for Strategy and Security at the Atlantic Council** [Tate and Margarita, 5-25-2022, “Eye to eye in AI: Developing artificial intelligence for national security and defense” Atlantic Council <https://www.atlanticcouncil.org/in-depth-research-reports/report/eye-to-eye-in-ai/> ARD]

Much of the urgency driving the DoD’s AI development and adoption efforts stems from the need to ensure the United States and its allies outpace China in the military-technological competition that has come to dominate the relationship between the two nations. Russia’s technological capabilities are far less developed, but its aggression undermines global security and threatens US and NATO interests. China China has prioritized investment in AI for both defense and national security as part of its efforts to become a “world class military” and to gain advantage in future “intelligentized” warfare—in which AI (alongside other emerging technologies) is more completely integrated into military systems and operations through “networked, intelligent, and autonomous systems and equipment.”21 While the full scope of China’s AI-related activities is not widely known, an October 2021 review of three hundred and forty-three AI-related Chinese military contracts by the Center for Security and Emerging Technology (CSET) estimates that PLA “spends more than $1.6 billion each year on AI-related systems and equipment.”22 The National Security Commission on Artificial Intelligence’s (NSCAI) final report assessed that “China’s plans, resources, and progress should concern all Americans. It is an AI peer in many areas and an AI leader in some applications.”23 CSET’s review and other open-source assessments reveal that China’s focus areas for AI development, like those of the United States, are broad, and include24 ● intelligent and autonomous vehicles, with a particular focus on swarming technologies; ● intelligence, surveillance, and reconnaissance (ISR); ● predictive maintenance and logistics; ● information, cyber, and electronic warfare; ● simulation and training (to include wargaming); ● command and control (C2); and ● automated target recognition. Progress in each of these areas constitutes a challenge to the United States’ capacity to keep pace in a military-technological competition with China. However, it is worth examining China’s advancing capabilities in two areas that could have a particularly potent effect on the military balance. Integration First, AI can help the PLA bridge gaps in operational readiness by artificially enhancing military integration and cross-domain operations. Many observers have pointed to the PLA’s lack of operational experience in conflict as a critical vulnerability. As impressive as China’s advancing military modernization has been from a technological perspective, none of the PLA’s personnel have been tested under fire in a high-end conflict in the same ways as the US military over the last twenty years. The PLA’s continuing efforts to increase its “jointness” from an organizational and doctrinal standpoint is also nascent and untested. The use of AI to improve the quality, fidelity, and complexity of simulations and wargames is one way the PLA is redressing this area of concern. A 2019 report by the Center for a New American Security observed that “[for] Chinese military strategists, among the lessons learned from AlphaGo’s victory was the fact that an AI could create tactics and stratagems superior to those of a human player in a game that can be compared to a wargame” that can more arduously test PLA decision-makers and improve upon command decision-making.25 In fact, the CSET report found that six percent of the three hundred and forty-three contracts surveyed were for the use of AI in simulation and training, including use of AI systems to wargame a Taiwan contingency.26 The focus on AI integration to reduce perceived vulnerabilities in experience also applies to operational and tactical training. In July 2021, the Chinese Communist Party mouthpiece publication Global Times reported that the PLA Air Force (PLAAF) has started to deploy AI as simulated opponents in pilots’ aerial combat training to “hone their decision-making and combat skills against fastcalculating computers.”27 Alongside virtual simulations, China is also aiming to use AI to support pilot training in real-world aircraft. In a China Central Television (CCTV) program that aired in November 2020, Zhang Hong, the chief designer of China’s L-15 trainer, noted that AI onboard training aircraft can “identify different habits each pilot has in flying. By managing them, we will let the pilots grow more safely and gain more combat capabilities in the future.”28 Notably, the PLAAF’s July 2021 AI–human dogfight was similar to the Defense Advanced Research Projects Agency’s (DARPA) September 2020 AlphaDogFight Challenge in which an AI agent defeated a human pilot in a series of five simulated dogfights.29 Similarly, the United States announced in September 2021 the award of a contract to training-and-simulation company Red 6 to integrate the company’s Airborne Tactical Augmented Reality System (ATARS)—which allows a pilot flying a realworld plane to train against AI-generated virtual aircraft using an augmented-reality headset—into the T-38 Talon trainer with plans to eventually install the system in fourth-generation aircraft.30 AI-enabled training and simulation are, therefore, key areas in which the US military is in a direct competition with the PLA. As the Chinese military is leveraging AI to enhance readiness, the DoD cannot afford to fall behind. Autonomy A second area of focus for Chinese AI development is in autonomous systems, especially swarming technologies, in which several systems will operate independently or in conjunction with one another to confuse and overwhelm opponent defensive systems. China’s interests in, and capacity for, developing swarm technologies has been well demonstrated, including the then record-setting launch of one hundred and eighteen small drones in a connected swarm in June 2017.31

**China is winning the AI arms race now – they are building capacity across their military spectrum.**

**Middendorf, 2021 - Former Secretary of the Navy** [J William “Opinion: Artificial Intelligence’s Military Risks, Potential” <https://www.govtech.com/news/opinion-artificial-intelligences-military-risks-potential.html>, BK]

With the emerging priority of artificial intelligence (AI), China is shifting away from a strategy of neutralizing or destroying an enemy’s conventional military assets — its planes, ships and army units. AI strategy is now evolving into dominating what are termed adversaries’ “systems-of-systems” — the combinations of all their intelligence and conventional military assets. What China would attempt first is to disable all of its adversaries’ information networks that bind their military systems and assets. It would destroy individual elements of these now-disaggregated forces, probably with missiles and naval strikes. Now, everything from submarines to satellites, tanks to jets, destroyers to drones, are AI connected by China. The People’s Liberation Army is developing autonomous vehicles that scout ahead of manned machines or provide supporting fire alongside them. These machines would be smart enough that a single human could supervise a whole pack of them. By replacing humans with electronics, combat vehicles will be more fuel-efficient, harder to hit and cheaper to build and operate. With AI at the helm, a central command could launch a multi-pronged attack from land, air and water simultaneously without any humans at the warfront.

### --Extend – Russia is Racing

**Russian AI will pose a major threat even if they are lagging in development**

**Nurkin and Konaev, 2022 - senior fellows at the Center for Strategy and Security at the Atlantic Council** [Tate and Margarita, 5-25-2022, “Eye to eye in AI: Developing artificial intelligence for national security and defense” Atlantic Council <https://www.atlanticcouncil.org/in-depth-research-reports/report/eye-to-eye-in-ai/> ARD]

Russia Russia lags behind the United States and China in terms of investments and capabilities in AI. The sanctions imposed over the war in Ukraine are also likely to take a massive toll on Russia’s science and An example of collaborative swarming drones on display at the UMEX 2022 exhibition in Abu Dhabi in February. Source: Tate Nurkin. 12 ATLANTIC COUNCIL technology sector. That said, US national decisionmakers should not discount Russia’s potential to use AI-enabled technologies in asymmetric ways to undermine US and NATO interests. The Russian Ministry of Defense has numerous autonomy and AIrelated programs at different stages of development and experimentation related to military robotics, unmanned systems, swarming technology, earlywarning and air-defense systems, ISR, C2, logistics, electronic warfare, and information operations.36 Russian military strategists see immense potential in greater autonomy and AI on future battlefields to speed up information processing, augment decision-making, enhance situational awareness, and safeguard the lives of Russian military personnel. The development and use of autonomous and AI-enabled systems are also discussed within the broader context of Russia’s military doctrine. Its doctrinal focus is on employing these technologies to disrupt and destroy the adversary’s command-and-control systems and communication capabilities and use non-military means to establish information superiority during the initial period of war, which, from Russia’s perspective, encompasses periods of non-kinetic conflict with adversaries like the United States and NATO.37

### AT Cyber Scenario

**No risk of collapse from disinformation – states will collaborate based on mutual fears of faulty AI and officers will discount false data**

**Fitzpatrick, 2019 - former the IISS Non-Proliferation and Nuclear Policy Programme** [Mark, 21 May, “Artificial Intelligence and Nuclear Command and Control,”  [https://www.iiss.org/blogs/survival-blog/2019/04/artificial-intelligence-nuclear-strategic-stability 6/18/22](https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/00396338.2019.1614782%206/18/22) MD]

Firstly, AI cannot be fully trusted. In particular, it can be risky to rely heavily on situational awareness generated by AI outputs. In the hypothetical scenario, both the US and Russian teams discounted the AI warnings that appeared inconsistent with human observations and were later found to have been generated by false data. This raises the possibility that resilient doubts about the reliability of AI could significantly attenuate its operational and strategic impact. Secondly, regarding emerging technology and nuclear strategy, the roll-out of new technologies – in the case at hand, sensor technology for detecting submarines – affects states differently depending on their strategic force structure. The Chinese team was most alarmed about the technology because, among the major players, it has the smallest submarine fleet and one that so far does not range far beyond China’s adjacent waters. Once identified, its submarines can thus be more easily neutralised. The exposure of China’s submarines before those of other players reinforced Beijing’s concerns that it was being targeted. Thirdly, shared concerns about AI-generated disinformation could foster collaboration among states to address the problem via confidence-building mechanisms. In the exercise, when the operating systems of Chinese nuclear power plants ‘raised the warning noise floor’ in ways that appeared intended to weaponise them, the US team shared information it had about similar noise warnings in American plants. The IAEA became the forum of choice for cooperative efforts towards a solution. More broadly, the potentially destabilising risks inherent in emerging technologies, as demonstrated in the Chinese nuclear-power example, could push states to proactively promote arms control. In our scenario, this prospect loomed only as the world stood on the brink of disaster. The takeaway is that policymakers can be educated in advance about such risks before they actually arise in a crisis and work to mitigate them.

### Solvency Resps

**1. TEVV fails for AI – because it is adaptive, it is impossible to predict situations to test, and it is too failure prone even in the best circumstances.**

**Lohn, 2020- Senior Fellow at Georgetown's Center for Security and Emerging Technology** [Andrew, 2 September, “Estimating The Brittleness Of Ai: Safety Integrity Levels And The Need For Testing Out-Of-Distribution Performance” [https://arxiv.org/abs/2009.00802 6/18/22](https://arxiv.org/abs/2009.00802%206/18/22) ST]

1 Introduction Test, Evaluation, Verification, and Validation (TEVV) for Artificial Intelligence (AI) is a central challenge that threatens to limit the economic and societal rewards that AI researchers have devoted themselves to producing. [1–5] TEVV for AI is particularly challenging for several reasons [6, 7] including that AI is meant to be used in circumstances that the designers cannot fully envision. The decision about what actions to take is made by the AI during use rather than by the designer before testing. The benefit is that during use, the AI will have access to the full range of inputs and environmental data to make decisions. The drawback is that it is too late for additional TEVV at that point. For example, non-AI software for autonomous braking would have well-delineated responses to the sensors such as radar and velocity. The decisions about how the vehicle will respond are made by the designer at design time and written into software which can have its code reviewed and tested to ensure that responses match the design decisions. For an AI, the designer has not made explicit decisions about how the vehicle will behave. Those decisions are made at use time, after testing and certification has already been completed. In cases where the range of inputs and environments that can be encountered are well understood, then testing is a straightforward matter. It may be arduous and expensive to gather enough test cases to certify the system, but it is conceptually simple to do. What makes TEVV for AI difficult is that often the range of inputs and environments that can be encountered is not well understood. The system is likely to encounter situations that are outside of the distributions of scenarios it was designed for. AI has been known to fail in those circumstances and has been commonly disparaged as brittle. [8–11] Brittleness implies two things about a component, first that it is highly functioning within some bounds and second, that it breaks readily when those bounds are exceeded. This report argues that neither of those criteria are as plain as often presented. First, compared to the reliability required of safety- or mission-critical systems for which TEVV and certification are paramount, the most highly touted AI successes are orders of magnitude more failure-prone even when being evaluated on data drawn from the same distributions they were designed for. And second, the performance of those models degrades smoothly as those bounds on the data distributions are relaxed, at a rate that is sometimes comparable to humans.

**2. No solvency – even if we make our AI safer, China and Russia won’t, so accidents and miscalculation will still happen. This specifically answers their authors**

**Allen, 2022 - director of the AI Governance Project at the CSIS** [Gregory, May 20, “One Key Challenge for Diplomacy on AI: China’s Military Does Not Want to Talk” [https://www.csis.org/analysis/one-key-challenge-diplomacy-ai-chinas-military-does-not-want-talk Acc 6/6/22](https://www.csis.org/analysis/one-key-challenge-diplomacy-ai-chinas-military-does-not-want-talk%20Acc%206/6/22) TA]

Over the past 10 years, artificial intelligence (AI) technology has become increasingly critical to scientific breakthroughs and technology innovation across an ever-widening set of fields, and warfare is no exception. In pursuit of new sources of competitive advantage, militaries around the world are working to accelerate the integration of AI technology into their capabilities and operations. However, the rise of military AI has brought with it fears of a new AI arms race and a potential new source of unintended conflict escalation. In the May/June 2022 issue of Foreign Affairs, Michael C. Horowitz, Lauren Kahn, and Laura Resnick Samotin write: “The United States, then, faces dueling risks from AI. If it moves too slowly, Washington could be overtaken by its competitors, jeopardizing national security. But if it moves too fast, it may compromise on safety and build AI systems that breed deadly accidents. Although the former is a larger risk than the latter, it is critical that the United States take safety concerns seriously.” Such fears are not entirely unfounded. Machine learning, the technology paradigm at the heart of the modern AI revolution, brings with it not only opportunities for radically improved performance, but also new failure modes. When it comes to traditional software, the U.S. military has decades of institutional muscle memory related to preventing technical accidents, but building machine learning systems that are reliable enough to be trusted in safety-critical or use-of-force applications is a newer challenge. To its credit, the Department of Defense (DOD) has devoted significant resources and attention to the problem: partnering with industry to make commercial AI test and evaluation capabilities more widely available, announcing AI ethics principles and releasing new guidelines and governance processes to ensure their robust implementation, updating longstanding DOD system safety standards to pay extra attention to machine learning failure modes, and funding a host of AI reliability and trustworthiness research efforts through organizations like the Defense Advanced Research Projects Agency (DARPA). However, even if the United States were somehow to successfully eliminate the risk of AI accidents in its own military systems—a bold and incredibly challenging goal, to be sure—it still would not have solved risks to the United States from technical failures in Russian and Chinese military AI systems. What if a Chinese AI-enabled early warning system erroneously announces that U.S. forces are launching a surprise attack? The resulting Chinese strike—wrongly believed to be a counterattack—could be the opening salvo of a new war.

**3. TEVV testing is Outdated--- current testing isn’t AI specific, conditions are hard to simulate, and failure rates probe the challenge**

**Lohn, 2020- Senior Fellow at Georgetown's Center for Security and Emerging Technology** [Andrew, 2 September, “Estimating The Brittleness Of Ai: Safety Integrity Levels And The Need For Testing Out-Of-Distribution Performance” [https://arxiv.org/abs/2009.00802 6/18/22](https://arxiv.org/abs/2009.00802%206/18/22) ST]

Conclusions AI is being considered for, or even applied in, critical industries. Those industries have well-established procedures and standards for incorporating new technologies but there are some mismatches between AI and those standards and procedures. First, AI commonly has failure rates that are orders of magnitude higher than the standards used in those critical industries. Second, AI is intended for tasks that do not accommodate straightforward evaluation. The central challenge facing the tester is to estimate the frequency and extent of departures from designed conditions that the AI will encounter in the real world and estimate the AI’s performance in those conditions. Fortunately, AI algorithms do maintain some level of performance outside of their design conditions such as on image classification of OOD inputs and perhaps speech recognition with varying accents or noise. If AI can reach the levels of performance in the best of conditions (perfectly in-distribution sampling) that are required for use in critical systems then evaluating performance in non-ideal conditions will be necessary. More work is needed to study the performance degradation of AI algorithms when subjected to OOD samples and to identify ways to improve their OOD performance but it is an area that holds some promise.

**4. DOD TEVV cannot solve for AI – the military branches are fragmented and there is no office to coordinate them.**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

BUREAUCRATIC BARRIERS In addition to these technological features, there are a number of bureaucratic barriers—ranging from leadership and process to human capital and infrastructure— preventing DOD from accelerating the development of new approaches to TEVV for ML/DL. Responsibility for ML/DL TEVV is shared and not well coordinated. While responsibility for TEVV is shared across multiple parts of the Office of the Secretary of Defense (OSD) and the services, greater coordination is needed to streamline investment and R&D on new testing approaches, increase cross-program visibility, and proliferate standards and best practices. There is a growing community of stakeholders within DOD and the broader U.S. government that will be critical to adapting the ML/DL TEVV enterprise. The Director of Operational Test & Evaluation (DOT&E) oversees policy and procedure for operational testing of major defense acquisition programs (MDAPs). DOT&E can play a key role in promulgating testing standards but tends to be cautious in setting new standards. It is accustomed to a rigid, sequential TEVV process that works well for MDAPs, but not for emerging technologies like ML/DL. The Testing Resource Management Center (TRMC) oversees infrastructure and spending, and develops investment roadmaps for new technology programs. TRMC will also be critical to adapting infrastructure for ML/DL DevSecOps. TRMC has included the Autonomy and Artificial Intelligence Test Technology Area in its T&E/S&T portfolio.13 The Joint Artificial Intelligence Center (JAIC), the Under Secretary of Defense (USD) for Research and Engineering (R&E), and the Director of the Defense Advanced Research Projects Agency (DARPA) all have important roles to play in the development of AI TEVV metrics, methods, and standards for DOD systems. The JAIC is actively engaged in setting standards, sharing best practices, and conducting testing. These programs promote, for example, designing DOD ML/DL systems and tagging data in ways that make it possible to understand how any particular decision is made. In April 2020, JAIC issued a request for information for new T&E capabilities for AI technologies.14 It is also already leading on implementation of the Defense Department’s AI ethics principles and integration of TEVV throughout the product development life cycle. The JAIC has established a DOD-wide responsible AI subcommittee, with representation from the services and the Joint Staff, DARPA, R&E, T&E, A&S, Policy, and the Office of the General Counsel to develop detailed policy documents, which will map the AI principles to the AI product life cycle and acquisition process.15 However, the JAIC is too small to scale these solutions throughout the Department. Meanwhile, the USD R&E is responsible for prototyping systems and developing large system of systems that will increasingly be AI-enabled. Finally, DARPA’s Explainable AI program is working to produce more explainable models that facilitate trust and human-machine collaboration. The armed services each have their own AI programs, which include testing components and research on AI TEVV at the service labs. The services have the operational knowledge and program acquisition offices and have traditionally led on developmental testing for major programs of record. They also know there is power in owning the test data and, understandably, want to evaluate the capabilities they are sending to their servicemembers themselves. However, the services don’t tend to have the S&T expertise and personnel to develop new approaches to TEVV for ML/DL. The Defense Department will need to designate an office or organization with overall responsibility for the TEVV process and establish a coordination mechanism that leverages the unique value-add of each of these entities, breaks down bureaucratic siloes, and streamlines investment in research and infrastructure to support new TEVV approaches. DOD policy, standards, and metrics for testing performance and evaluating risk need to evolve. DOD needs a policy framework for determining safety standards for a range of ML/DL applications based on the use case, mission, and anticipated environment in which the system will operate. These standards then need to be translated into requirements for system design and metrics for measuring system performance that are operationally-relevant; transparent to developers, testers, and users; and reflect DoD’s AI ethics and U.S. values.

**5. TEVV fails due to DOD bureaucracy, insufficient infrastructure and tests that are inappropriate for AI – plan doesn’t change Any of these**

**Nurkin and Konaev, 2022 - senior fellows at the Center for Strategy and Security at the Atlantic Council** [Tate and Margarita, 5-25-2022, “Eye to eye in AI: Developing artificial intelligence for national security and defense” Atlantic Council <https://www.atlanticcouncil.org/in-depth-research-reports/report/eye-to-eye-in-ai/> ARD]

There are no easy solutions to this challenge. But, a collaborative process that engages stakeholders across government, industry, academia, and civil society could help prevent AI development from going down the path of social media, where public policy failed to anticipate and was slow to respond to the risks and damages caused by disinformation and other malicious activity on these platforms. Related to standards are the challenges linked to testing, evaluation, verification, and validation (TEVV). Testing and verification processes are meant to “help decision-makers and operators understand and manage the risks of developing, producing, operating, and sustaining AI-enabling systems,” and are essential for building trust in AI.81 The DoD’s current TEVV protocols and infrastructure are meant primarily for major defense acquisition programs like ships, airplanes, or tanks; it is linear, sequential, and, ultimately, finite once the program transitions to production and deployment. With AI systems, however, “development is never really finished, so neither is testing.”82 Adaptive, continuously learning emerging technologies like AI, therefore, require a more agile and iterative development-and-testing approach—one that, as the NSCAI recommended, “integrates testing as a continuous part of requirements specification, development, deployment, training, and maintenance and includes run-time monitoring of operational behavior.”83 An integrated and automated approach to development and testing, which builds upon the commercial best practice of development, security, and operations (DevSecOps), is much better suited for AI/ML systems. While the JAIC’s JCF has the potential to enable a true AI DevSecOps approach, scaling such efforts across the DoD is a major challenge because it requires significant changes to the current testing infrastructure, as well as more resources such as bandwidth, computing support, and technical personnel. That said, failing to develop new testing methods better suited to AI, and not adapting the current testing infrastructure to support iterative testing, will stymie efforts to integrate and adopt trusted and responsible AI at scale. The above discussion of standards and TEVV encapsulates the unique challenges modern AI systems pose to existing DoD frameworks and processes, as well as the divergent approaches commercial technology companies and the DoD take to AI development, deployment, use, and maintenance. To accelerate AI adoption, the DoD and its industry partners need to better align on concrete, realistic, operationally relevant standards and performance requirements, testing processes, and evaluation metrics that incorporate ethical AI principles. A defense-technology ecosystem oriented around trusted and responsible AI could promote the cross-pollination of best practices and lower the bureaucratic and procedural barriers faced by nontraditional vendors and startups.

**6. Even with testing, Safety is Impossible – lack of critical human understandings leads to instability and accidents**

**Horowitz and Scharre 2021 - Director of the Emerging Capabilities Policy Office in the Office of the Under Secretary of Defense for Policy and  Vice President and Director of Studies at CNAS** [Michael, Paul, January 12 2021, “AI and International Stability: Risks and Confidence- Building Measures”, <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>, Acc 6/18/22. M.A.]

Even if military AI systems are adequately tested, the use of AI to enable more autonomous machine behavior in military systems raises an additional set of risks. In delegating decision-making from humans to machines, policymakers may de facto be fielding forces with less flexibility and ability to understand context, which would then have deleterious effects on crisis stability and managing escalation. While machines have many advantages in speed, precision, and repeatable actions, machines today cannot come close to human intelligence in understanding context and flexibly adapting to novel situations. This brittleness of machine decision-making may particularly be a challenge in pre-conflict crisis situations, where tensions among nations run high. Military forces from competing nations regularly interact in militarized disputes below the threshold of war in a variety of contested regions (e.g., the India-Pakistan border, China-India border, South China Sea, Black Sea, Syria, etc.). These interactions among deployed forces sometimes run the risk of escalation due to incidents or skirmishes that can inflame tensions on all sides. This poses a challenge for national leaders, who have imperfect command-and-control over their own military forces. Today, however, deployed military forces rely on human decision-making. Humans can understand broad guidance from their national leadership and commander’s intent, such as “defend our territorial claims, but don’t start a war.” Relative to humans, even the most advanced AI systems today have no ability to understand broad guidance, nor do they exhibit the kinds of contextual understanding that humans frequently label “common sense.”[27](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn27) Militaries already employ uninhabited vehicles (drones) in contested areas, which have been involved in a number of escalatory incidents in the East China Sea, South China Sea, Syria, and Strait of Hormuz.[28](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn28) Over time, as militaries incorporate more autonomous functionality into uninhabited vehicles, that functionality could complicate interactions in these and other contested areas. Autonomous systems may take actions based on programming that, while not a malfunction, are other than what a commander would have wanted a similarly situated human to do in the same situation. While the degree of flexibility afforded subordinates varies considerably by military culture and doctrine, humans have a greater ability to flexibly respond to complex and potentially ambiguous escalatory incidents in ways that may balance competing demands of ensuring national resolve while managing escalation.[29](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn29) Autonomous systems will simply follow their programming, whatever that may be, even if those rules no longer make sense or are inconsistent with a commander’s intent in the given situation. This challenge is compounded by the fact that human commanders cannot anticipate all of the possible situations that forward-deployed military forces in contested regions may face. Employing autonomous systems in a crisis effectively forces human decision makers to tie their own hands with certain pre-specified actions, even if they would rather not.

**7. AI TEVV does not solve safety – tests are unreliable, it causes huge time delays, and has high failure rates.**

**Lohn, 2020- Senior Fellow at Georgetown's Center for Security and Emerging Technology** [Andrew, 2 September, “Estimating The Brittleness Of Ai: Safety Integrity Levels And The Need For Testing Out-Of-Distribution Performance” [https://arxiv.org/abs/2009.00802 6/18/22](https://arxiv.org/abs/2009.00802%206/18/22) ST]

4 SILs for AI With a rough intuition for the safety goals to achieve, one can begin to assess the feasibility of developing various AI technologies that reach the required reliability as well as developing the testing methods that would be needed to certify them. 4.1 Reliability in AI AI is typically tasked with difficult problems. For example, performance on Image-Net classification has been one of AI’s most lauded successes. At the time of writing, it is led by groups posting top-1 accuracies in the upper 80’s of percent and top-5 accuracies in the upper 90’s [32] [33] [34] and disease detection from medical imaging is comparable. [35] That incredible level of performance is the result of the combined efforts of a massive number of researchers over many years but still the failure rates are greater than 10−2/use. Those failure rates correspond to the lowest SIL rating in all of Table 2 for any industry. When considered in a continuous setting where many images are evaluated per hour, the situation is far worse. In that case, the failure rates are too high for any SIL. To reach even the minimum continuous-demand level would require more than ten hours between attempts for top-5 images and more than 100 hours between attempts for top-1 images. Those attempt frequencies are far from what most AI practitioners consider continuous. Processing even just ten images per second would require an accuracy of 0.99999997 to get to even the lowest level in aviation 10−3/h, and the lowest level for the automotive industry is three orders of magnitude more restrictive than that. Viewed in this light, AI is not confronted with a testing challenge, it is facing a reliability challenge. Getting to one-in-a-million failure rates in image classification for diverse sets of objects is perhaps an unfair goal. Image-Net classifiers already outperform humans. That said, critical systems and their certification procedures have been designed over a century to accommodate humans as a fallible component of a reasonably reliable system. Further, as mentioned earlier, certification of AI as an operator rather than as software does not appear to be a near-term path forward. Either way, the message to be received by AI designers and practitioners is that the failure rates of the most heralded algorithms are orders of magnitude more failure-prone than safety-critical systems typically certify.

**8. Safety testing for AI is difficult because AI is Opaque – even certifying boards admit it.**

**Lohn, 2020- Senior Fellow at Georgetown's Center for Security and Emerging Technology** [Andrew, 2 September, “Estimating The Brittleness Of Ai: Safety Integrity Levels And The Need For Testing Out-Of-Distribution Performance” [https://arxiv.org/abs/2009.00802 6/18/22](https://arxiv.org/abs/2009.00802%206/18/22) ST]

2.1 Processes and Standards are Complex and Opaque AI is a ubiquitous technology that can be envisioned in an infinity of applications. Many of those applications warrant deeper levels of scrutiny than others because of heightened risks and, for the most part, those applications have already evolved processes for obtaining high levels of surety. Those processes tend to be complex and opaque to an extent where people build entire careers around understanding and navigating them. Additionally, most of the standards documents (the notable exception being military standards) are behind paywalls, restricting access and widespread understanding. As a result of the opacity, complexity, and diversity of these processes, it is probably not feasible for most AI designers to be well-versed in even one of them let alone the wide range of applications for which a single model or architecture may be used. Nevertheless, currently much of the onus for AI safety, and much of the blame for failures, falls on the AI designers. While recognizing that the enormity of the problem will require many dedicated staff whose sole job is translating safety goals into design specifications, all involved should have a basic understanding of the processes, requirements, and goals. The next few subsections will very briefly highlight a few of the takeaways that are relevant to AI designers and practitioners from a survey of the processes for certifying aircraft, nuclear power plants, automobiles, pharmaceuticals, and weapons systems. Their processes and standards are far from perfect as is recognized by the certifying bodies themselves who have sponsored studies [12, 13] to search out better alternatives. There have been major and minor revisions to the standards but the existing approaches seem to hold up well to scrutiny as compared to other options that have been proposed.

**9. The JAIC Fails on TEVV because it lacks visibility and clarity**

**Tarraf, Shelton, and Parker 2019 - Senior Information Scientist, Senior Engineer, and physical Scientist at the RAND Corporation** [ Danielle C., William, Edward, et al RAND Cooperation, “The Department of Defense Posture for Artificial Intelligence”, <file:///Users/MiraAgarwal/Downloads/RAND_RR4229%20(2).pdf>, Acc 6/18/22, M.A.]

The JAIC lacks visibility. The issue of visibility is subtle. The JAIC has been designated, in the summary of the DoD AI strategy, as the focal point for carrying out DoD’s strategy, and is expected to attract and cultivate a world-class AI team. This designation and role presume a certain degree of visibility both within DoD and outside it. This visibility was lacking based on our interviews. Overall, we noted a lack of clarity among our interviewees on the JAIC’s mandate, roles, and activities. We also noted a lack of clarity around how it fits within the broader DoD ecosystem and how it connects to the services and their efforts. That was true of both DoD interviewees and our industry interviewees who had heard of the JAIC (see section “Organization: At the OSD Level” in Appendix B and section “Thoughts Across Industry: On the JAIC” in Appendix C). In addition to the lack of clarity about the JAIC’s current mandate and roles, there were many perspectives about the desired or ideal role for the JAIC as DoD embraces and scales AI. These perspectives ranged from the JAIC as a central repository of information and best practices, to the JAIC as a center of excellence that focuses on discrete tasks (e.g., building the JCF, formalizing standards for VVT&E), to the JAIC’s potential elevation to a field agency or other entity with a direct reporting line to either the Secretary or DSD. We do not believe this lack of clarity to be simply a question of messaging. More fundamentally, it points to a lack of clarity about the raison d’être of the JAIC and how the specific roles it has been assigned support that. The confusion might not be entirely on the part of the audience. DoD needs to have a clearer view of what it wants the JAIC to be and how DoD can help ensure the success of JAIC’s mission, and therefore DoD’s vision.

**10. TEVV fails because of a lack of Training Data – the aff does not correct this problem.**

**Atherton 2022 – Military Technology Journalist** [Kelsey, 5/6/22, “Understanding the errors introduced by military AI applications”, <https://www.brookings.edu/techstream/understanding-the-errors-introduced-by-military-ai-applications/>, 6/18/22, LND]

This dynamic played out in the other fatal friendly fire incident involving a Patriot missile battery during the Iraq War, when a U.S. Navy F/A-18 aircraft was misidentified as a ballistic missile and shot down, killing the pilot. According to a 2019 Center for Naval Analyses [report](https://www.cna.org/CNA_files/PDF/DOP-2019-U-021957-1Rev.pdf), the Patriot recommended that the operator fire missiles in response to what it had identified as an enemy projectile, and the operator approved the recommendation to fire “without independent scrutiny of the information available to him.” This difficulty faced by Patriot missile batteries in correctly identifying potential targets illustrates one of the most serious challenges facing autonomous weapons—getting accurate training data. As militaries move toward greater autonomy in a wide range of systems, they are increasingly reliant on machine learning technology that uses large data sets to make predictions about how a machine should operate. The challenge of acquiring accurate data sets autonomous systems up for inevitable failure. “Conflict environments are harsh, dynamic and adversarial, and there will always be more variability in the real-world data of the battlefield than in the limited sample of data on which autonomous systems are built and verified,” as Arthur Holland Michel, and associate researcher in the Security and Technology Programme at the UN Institute for Disarmament Research, wrote in [a report](https://unidir.org/publication/known-unknowns) last year addressing data issues in military autonomous weapons. A lack of reliable data or an inability to produce datasets that replicate combat conditions will make it more likely that autonomous weapons fail to make accurate identifications.

**11. TEVV fails – there is a lack of a trained AI Workforce and recruitment is minimal**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

DoD lacks the ability to recruit, train, and retain the right talent. For many organizations within the DoD TEVV ecosystem, recruiting and retaining talent is often a bigger challenge than securing funding. These organizations need diverse, interdisciplinary teams that understand both testing and the technology itself. DoD needs data scientists, statisticians, and computer scientists that can develop new testing and verification mechanisms; computer science and ML/DL experts to develop the technology; and operators that understand the technology enough to trust, deploy, and integrate it operationally. Finally, it needs experts in human cognition and psychology that understand human-machine interaction and can build interfaces that enable greater trust. Many of the challenges of recruiting and retaining such technical talent are not unique. Existing DoD programs to recruit recent science, technology, engineering, and math graduates are too small, non-traditional hiring authorities for STEM talent are underutilized, and the service academies do not feed enough STEM talent directly into technical roles. DoD lacks dedicated career paths for technologists and testers, which further constrains the Department’s ability to retain what talent it does manage to recruit or grow in-house. Not all of this talent needs to be cutting-edge researchers; the Department will need a cadre of professionals—program managers, requirements writers, lawyers, operators, policy officials, and others—who have a baseline understanding of the technology and testing procedures, and can bridge the gap between DoD leadership and policy teams on one hand and the technical developers and testers on the other. Further, DoD should leverage its expansive network of FFRDCs and academic partnerships to expand its access to technical personnel. Many of the FFRDCs, such as the Lawrence Berkeley and Lawrence Livermore National Laboratories in California and the MIT Lincoln Laboratory in Cambridge, are located near hotspots for AI talent and have fewer challenges with hiring.

**12. It is impossible to regulate AI without a clear Definition – that is necessary to governing this technology**

**Artificial Intelligence/Machine Learning Risk & Security Working Group, 2020** [Wharton, ‘Artificial Intelligence Risk & Governance’, <https://ai.wharton.upenn.edu/artificial-intelligence-risk-governance/>]

3.2.1 Definitions Depending on the adoption, environment, and culture of an organization, there may be a long series of nuanced definitions for AI/ML. As the first step to achieving AI governance, a clear definition of what constitutes AI (and what does not) is critical for any organization. This definition provides the foundation and establishes a clear understanding of the other components of the governance structure, informing the remaining building blocks that comprise the overall AI program (e.g., inventory). Any definition of AI should consider, among other factors, the variety of techniques used by the organization in training and developing the AI, what distinguishes AI from other traditional rule-based systems, and the implications of the definition, enabling the kind of AI inventory efforts we set forth below. Definitions and supporting documentation should provide clarity related to how various stakeholders – including Senior Management, Legal, System Developers, Compliance, and Information Security Officers – identify with the AI definition relative to other well-established definitions.[6]

**13. TEVV is inadequate for AI – Neural networks are far too failure prone, and performance falls off.**

**Lohn, 2020- Senior Fellow at Georgetown's Center for Security and Emerging Technology** [Andrew, 2 September, “Estimating The Brittleness Of Ai: Safety Integrity Levels And The Need For Testing Out-Of-Distribution Performance” [https://arxiv.org/abs/2009.00802 6/18/22](https://arxiv.org/abs/2009.00802%206/18/22) ST]

Test, Evaluation, Verification, and Validation (TEVV) for Artificial Intelligence (AI) is a challenge that threatens to limit the economic and societal rewards that AI researchers have devoted themselves to producing. A central task of TEVV for AI is estimating brittleness, where brittleness implies that the system functions well within some bounds and poorly outside of those bounds. This paper argues that neither of those criteria are certain of Deep Neural Networks. First, highly touted AI successes (eg. image classification and speech recognition) are orders of magnitude more failure-prone than are typically certified in critical systems even within design bounds (perfectly in-distribution sampling). Second, performance falls off only gradually as inputs become further Out-Of-Distribution (OOD). Enhanced emphasis is needed on designing systems that are resilient despite failure-prone AI components as well as on evaluating and improving OOD performance in order to get AI to where it can clear the challenging hurdles of TEVV and certification.

**14. Predicting AI is too difficult – it is under researched, misunderstood, and the military in inherently uncertain**

**Goldfarb and Lindsay, 2022 – Chair in AI and health care at the Univ of Toronto, Prof of Cybersecurity at the Georgia Tech** [Avi, Jon, 2/25/22, <https://doi.org/10.1162/isec_a_00425>, “Prediction and Judgement: Why Artificial Intelligence Increases the Importance of Humans in War” 6/18/22, LND]

Many hopes and fears about AI recapitulate earlier ideas about the information technology revolution in military affairs (RMA) and cyberwarfare.[9](javascript:;) Familiar tropes abound regarding the transformative effects of commercial innovation, the speed and danger of networked computation, the dominance of offense over defense, and the advantages of a rising China over a vulnerable United States. But skeptics have systematically challenged both the logic and empirical basis of these assumptions about the RMA[10](javascript:;) and cyberwar.[11](javascript:;) Superficially plausible arguments about information technology tend to ignore important organizational and strategic factors that shape the adoption and use of digital systems. As in the economics literature, an overarching theme in scholarship on military innovation is that technology is not a simple substitute for military power.[12](javascript:;) Technological capabilities depend on complementary institutions, skills, and doctrines. Furthermore, implementation is usually marked by friction, unintended consequences, and disappointed expectations. The RMA and cyber debates thus offer a cautionary tale for claims about AI. It is reasonable to expect organizational and strategic context to condition the performance of automated systems, as with any other information technology.[13](javascript:;) AI may seem different, nevertheless, because human agency is at stake. Recent scholarship raises a host of questions about the prospect of automated decision-making. How will war “at machine speed” transform the offense-defense balance?[14](javascript:;) Will AI undermine deterrence and strategic stability,[15](javascript:;) or violate human rights?[16](javascript:;) How will nations and coalitions maintain control of automated warriors?[17](javascript:;) Does AI shift the balance of power from incumbents to challengers or from democracies to autocracies?[18](javascript:;) These questions focus on the substitutes for AI because they address the political, operational, and moral consequences of replacing people, machines, and processes with automated systems. The literature on military AI has focused less on the complements of AI, namely the organizational infrastructure, human skills, doctrinal concepts, and command relationships that are needed to harness the advantages and mitigate the risks of automated decision-making.[19](javascript:;) In this article, we challenge the assumptions behind AI substitution and explore the implications of AI complements. An army of lethal autonomous weapon systems may well be destabilizing, and such an army may well be attractive to democracies and autocracies alike. The idea that machines will replace warriors, however, represents a misunderstanding about what warriors actually do. We suggest that it is premature to forecast radical strategic consequences without first clarifying the problem that AI is supposed to solve. We provide a framework that explains how the complements of AI (i.e., data and judgment) affect decision-making. In general, automation is advantageous when quality data can be combined with clear judgments. But the consummate military tasks of command, fire, and maneuver are fraught with uncertainty and confusion. In contrast, more institutionalized tasks in administration and logistics tend to have copious data and clear goals, which are conducive to automation. We argue that militaries risk facing bad or tragic outcomes if they conflate these conditions by prematurely providing autonomous systems with clear objectives in uncertain circumstances. Conversely, for intelligence and operational tasks that have quality data but difficult judgments, teams of humans and machines can distribute the cognitive load of decision-making. We expect many if not most military AI tasks to fall into the latter category, which we describe as human-machine teaming. The net result, we argue, is that data and judgment will become increasingly valuable and contested, and thus AI-enabled warfare will tend to become more protracted and confusing.

**15. No Industry Cooperation - Lack of formalized communication between AI builders and users limits openness and trust necessary for TEVV**

**Tarraf, Shelton, Parker 2019 - Senior Information Scientist, Senior Engineer, and physical Scientist at the RAND Corporation** [ Danielle C., William, Edward, et al RAND Cooperation, “The Department of Defense Posture for Artificial Intelligence”, <file:///Users/MiraAgarwal/Downloads/RAND_RR4229%20(2).pdf>, Acc 6/18/22, M.A.]

Communication channels among the builders—and users— of AI within DoD are sparse. For example, one of the takeaways from our interviews is that communication among the research organizations appears to be limited, and when it does occur, it is driven primarily by personal connections among program managers or researchers (see section “Advancement and Adoption” in Appendix B). This sparsity of communication is inconsistent with the culture of openness and sharing that was emphasized by our academic and industry interviewees as a driver of success (see section “Industry: Innovation” in Appendix C, and section “Academia: Advancement and Adoption” in Appendix C).24 Likewise, we noted AI RDT&E activities throughout the services, but our takeaway from the interviews was that visibility into these activities is limited, both within and across the services and from OSD. Finally, mechanisms of interactions between the developers 23 Isaac R. Porche, III, Shawn McKay, Megan McKernan, Robert Warren Button, Bob Murphy, Katheryn Giglio, and Elliot Axelband, Rapid Acquisition and Fielding for Information Assurance and Cyber Security in the Navy, Santa Monica, Calif.: RAND Corporation, TR-1294-NAVY, 2012. 24 We should emphasize here that the sparsity of communication appeared to be driven by the lack of formalized communication channels rather than an unwillingness to communicate. 54 The Department of Defense Posture for Artificial Intelligence (e.g., research entities) and users (e.g., warfighters, analytics officers) of AI are limited or nonexistent.25 There are many potential impediments to users adopting AI technologies. Those include an inherent resistance to change— including in roles and TTPs; concerns about the potential loss of an individual’s value to the organization as a result of the adoption of AI capabilities; and lacking trust in the technologies.26 These perceived impediments are not unique to DoD; our interviews in industry and academia highlighted similar concerns (Appendix C). Nonetheless, these are serious concerns, and ones that DoD needs to address to effectively scale AI. There is a lack of consensus on the delineation of AI investments within DoD. This finding points to a set of practical questions that DoD needs to answer: For the purpose of accounting for AI investments, what counts as an AI activity and what does not? As is also the case with software, DoD budgets do not account for AI when it is a small part of a larger platform, making it hard to track overall spending on AI. We note here that adopting a DoD-wide definition of AI does not necessarily provide an answer to these practical problems.27

**16, TEVV for AI fails, because it treats AI systems like Software, but when they are autonomous, they should be tested like Operators.**

**Lohn, 2020- Senior Fellow at Georgetown's Center for Security and Emerging Technology** [Andrew, 2 September, “Estimating The Brittleness Of Ai: Safety Integrity Levels And The Need For Testing Out-Of-Distribution Performance” [https://arxiv.org/abs/2009.00802 6/18/22](https://arxiv.org/abs/2009.00802%206/18/22) ST]

2.2 Is AI Software or an Operator The certification of new critical systems (such as a new aircraft or a power plant design) moves along established processes and standards which are each composed of many lower level standards and processes. The chains can be extensive. Among those lower level standards can be those dedicated to software components and those dedicated to operators. Although AI is clearly software (or perhaps electronic hardware), depending on the intended application it might be used to perform tasks that are more commonly associated with operators. Certification of operators tends to rely on passing tests, accumulating hours of experience, and continually monitoring fitness for duty. A driver’s license is a familiar example that requires a written and practical exam, starts with a permit, and can be revoked for poor performance or degraded operational capacity. More stringent examples are pilots [14] and nuclear operators [15] which require minimum education and minimum hours of supervised operation. The processes and standards that exist for operators were designed for humans and are not easily applied to AI so even in cases where the tasks being performed are more commonly associated with those of an operator than of traditional software, the processes and standards for software certification tend to be more appropriate for AI. Testing and certification that licenses AI as more of an operator is a promising direction of study [1] but does not appear to be a near-term solution. An important distinction to consider in operator-like approaches vs software-like approaches is that the diversity of human operators provides some assurance against systemic failures across the entire fleet but also precludes in-depth testing of the entire set of operators in a cost and time-efficient manner. That systemic risk and opportunity for in-depth testing is a main reason why software standards and processes are currently more applicable to AI.

#### 17. TEVV cannot solve – normal testing methods don’t include run time monitoring which is key to AI

Haugh, 2018 - project leader for the Institute for Defense Analysis [Brian A., September “The Status of Test, Evaluation, Verification, and Validation (TEV&V) of Autonomous Systems,” <https://www.ida.org/-/media/feature/publications/t/th/the-status-of-test-evaluation-verification-and-validation-of-autonomous-systems/p-9292.ashx>, 6/23/22 MD]

D. Run-Time Monitoring Given the difficulty of assuring that a system will not exhibit specific undesired behaviors, a natural thought is to instead monitor the system during operations and intervene when bad behavior is imminent. This approach is already common in engineering practice for safety-critical systems. It is mentioned explicitly in the recent US Unmanned Systems Integrated Roadmap 2017-2042: For the most demanding adaptive and non-deterministic systems, a new approach to traditional TEVV will be needed. For these types of highly complex autonomous systems, an alternate method leveraging a run-time architecture that can constrain the system to a set of allowable, predictable, and recoverable behaviors should be integrated early into the development process. Emergent behaviors from large-scale deployment of interacting autonomous systems poses a difficult challenge. The analysis and test burden would thereby, be shifted to a simpler, more deterministic run-time assurance mechanism. The effort for new approaches to TEVV endeavors to provide a structured argument, supported by evidence, justifying that a 3-6 system is acceptably safe and secure not only through offline tests, but also through reliance on real-time monitoring, prediction, and fail-safe recovery.1 Although this mechanism might indeed be simpler than avoiding unpredictable behaviors in the first place, it is not without its own challenges. In general, any behavior whose dependability cannot be adequately assured through system design and training would need to be monitored, with a robust intervention standing by. This means not only intervening when the system is about to execute some undesired physical action (e.g., one that might risk harm to the system or to humans), but also intervening in any case where the system is making a bad decision or misinterpreting its environment. Detecting such cases and handling them gracefully will not always be easy. Research is required into architectures to support this concept, instrumentation needs and control algorithms to predict and avoid specific failure modes, systematic identification of conditions to be monitored for, robustness against attacks designed to invoke fail-safe behaviors, and so forth. It goes without saying that the fail-safe systems would themselves need to be verified and validated as well.

### --Extend – Adaptive

**Testing and Validation fails for AI – machine learning means the system is constantly “reprogrammed” so initial testing becomes irrelevant**

**Artificial Intelligence/Machine Learning Risk & Security Working Group, 2020** [Wharton, ‘Artificial Intelligence Risk & Governance’, <https://ai.wharton.upenn.edu/artificial-intelligence-risk-governance/>]

2.1.3 Testing and Trust Depending on the implementation and use case, the AI system could potentially evolve over time at varying degrees. Some forms of AI could generate complexities that may accrue, evolve or worsen over time.[4] ML models may be sensitive to environmental developments, for example, that could potentially alter their performance, Some AI systems may not have exposures to the below potential risks, either due to the nature of implementation or controls in place. Potential concerns related to testing and trust risk are discussed in detail below: Incorrect Output Testing and validation of AI/ML systems may pose challenges relative to traditional systems as certain AI/ML systems are inherently dynamic, apt to change over time, and by extension, may result in changes to their outputs. Testing for all scenarios, permutations and combinations of available data may not be possible, thus leading to potential gaps in coverage. The severity of these gaps may vary with each system and its applications. Lack of Transparency As an emerging technology, the awareness of (and hype related to) AI and the lack of adequate understanding of the technology could potentially give rise to trust issues with AI systems. There is a perception, for example, that AI systems are a “black box” and therefore cannot be explained. (We address this belief further in Section 4.) Generally, it is difficult to thoroughly assess systems that cannot easily be understood. Bias AI systems could potentially amplify risks relating to unfairly biased outcomes or discrimination. For example, the subjects of data ethics, fairness and the possibility of unfairly biased outcomes from the use of AI are still evolving. It is evident, however, that, depending on the use case, there is a risk that AI systems could potentially lead to unfairly biased outcomes for individuals and/or organizations. Furthermore, AI-driven unfairly biased outcomes could have privacy compliance implications, constitute regulatory, litigation and reputational risk, impact operations and result in customer dissatisfaction and attrition. Section 4 of this paper (focused on transparency, explainability and bias) discusses unfairly biased outcomes and discrimination in AI in greater detail. 2.1.4 Compliance Policy Non-Compliance As AI implementations mature in organizations, their impact on existing internal policies should be considered. Regulatory bodies have expressed growing interest in AI deployments in the financial industry. Regulators have formed working groups representing various authorities across the globe to discuss supervisory challenges posed by emerging technologies, which have led to the publication of guidelines, white papers, and surveys. This interest is driven by the understanding that AI/ML poses new challenges, and readers should evaluate how regulations may impact the use and governance of AI/ML. AIRS is not advocating for new regulation(s), but merely would encourage readers to monitor existing regulations and their potential applicability to AI.

#### No Solvency - TEVV does not pair well with formal methods of research

Haugh, 2018 - project leader for the Institute for Defense Analysis [Brian A., September “The Status of Test, Evaluation, Verification, and Validation (TEV&V) of Autonomous Systems,” <https://www.ida.org/-/media/feature/publications/t/th/the-status-of-test-evaluation-verification-and-validation-of-autonomous-systems/p-9292.ashx>, 6/23/22 MD]

3. Potential Approaches to Overcome These Challenges Academic and government researchers have begun work on a number of techniques and methodologies that might help to overcome these core challenges. Current efforts fall into four primary categories: formal methods, cognitive instrumentation, adversarial testing, and run-time monitoring. We discuss each of these below, then consider the question of how well they cover the full set of capability gaps implied by the challenge list. A. Formal Methods 1. Summary Formal methods in software development allow developers to specify certain properties that the software should have, produce the software, and verify that it does have those properties without needing to confirm that empirically by testing for them. Properties to be specified might be things like: Property 1: the weapon cannot fire while turret is still rotating Property 2: the course-of-action selector can never get into an infinite loop There are two approaches to formal methods for autonomous systems verification: (1) formal methods can be used after the fact as an analytic tool to verify some properties of existing software, or more importantly, (2) formal methods can be used as a design and development process that can assure much more about the behavior of the software to be developed. Formal methods of the second kind are most commonly used in the development of complex safety-critical or security-critical systems or for expensive one-time development efforts (e.g., deep space probes). Applying formal methods to complex AI and autonomous systems is a natural extension of this. 2. Limitation Although formal methods can be extremely useful, there are significant constraints on the current state of the art. These include the following: Scalability: There are currently fairly tight bounds on the size of development effort (or state space) that the techniques can be applied to. 3-2 Scope: Not all desired properties can be assured through formal methods, and there may be performance trades associated with achieving assurance. Rigidity: Any change to a system developed using formal methods risks invalidating the assurance proofs, unless the formal methods are reapplied to the new specification. Given that not all desired behaviors can be assured using formal methods, there are also open research questions concerning how to combine formal methods with empirical TEV&V techniques or run-time monitoring strategies.

### --Extend – China and Russia

**China takes out solvency – they will do nothing to reduce accidents**

**Allen, 2022 - director of the AI Governance Project at the CSIS** [Gregory, May 20, “One Key Challenge for Diplomacy on AI: China’s Military Does Not Want to Talk” [https://www.csis.org/analysis/one-key-challenge-diplomacy-ai-chinas-military-does-not-want-talk Acc 6/6/22](https://www.csis.org/analysis/one-key-challenge-diplomacy-ai-chinas-military-does-not-want-talk%20Acc%206/6/22) TA]

The truth, unfortunately, is that—despite the United States’ efforts at transparency and requests for dialogue—the United States knows very little about how seriously the Chinese military considers ethics in its use of AI, how robust Chinese test and evaluation processes are, and what governance structures and procedures exist to reduce the risk of military AI accidents. That secrecy in and of itself is a source of risk to international peace and security. But, then again, what incentive does China have to substantively engage? The United States is already providing a great deal of transparency around its own risk reduction efforts, and China is already garnering many reputational benefits from calling for dialogue without any of the costs of substantively participating. Perhaps neither the U.S. government nor the Chinese scholarly community can succeed in persuading the PLA that it is in everyone’s best interest for this dialogue to occur. At the very least, however, it should be clear to the international community that China is the one refusing to talk.

**Chinese motives on Testing and Evaluation in AI are unknown due to secrecy.**

**Allen, 2022 - director of the AI Governance Project at the CSIS** [Gregory, May 20, “One Key Challenge for Diplomacy on AI: China’s Military Does Not Want to Talk” [https://www.csis.org/analysis/one-key-challenge-diplomacy-ai-chinas-military-does-not-want-talk Acc 6/6/22](https://www.csis.org/analysis/one-key-challenge-diplomacy-ai-chinas-military-does-not-want-talk%20Acc%206/6/22) TA]

The truth, unfortunately, is that—despite the United States’ efforts at transparency and requests for dialogue—the United States knows very little about how seriously the Chinese military considers ethics in its use of AI, how robust Chinese test and evaluation processes are, and what governance structures and procedures exist to reduce the risk of military AI accidents. That secrecy in and of itself is a source of risk to international peace and security. But, then again, what incentive does China have to substantively engage? The United States is already providing a great deal of transparency around its own risk reduction efforts, and China is already garnering many reputational benefits from calling for dialogue without any of the costs of substantively participating. Perhaps neither the U.S. government nor the Chinese scholarly community can succeed in persuading the PLA that it is in everyone’s best interest for this dialogue to occur. At the very least, however, it should be clear to the international community that China is the one refusing to talk.

### --Extend – Outdated

**TEVV fails for AI and machine learning – it is too rigid and sequential – Aff does not change the current methods**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

The current rigid, sequential development and testing process for major defense acquisition programs—such as hardware-intensive systems like ships, airplanes, or tanks—is not well suited for adaptive emerging technologies like ML/DL. The current technology acquisition process takes a linear, waterfall approach to development and testing. Companies must pass through a series of acquisition phases and milestone decision points—moving from prototyping/technology maturation to manufacturing and development to production and deployment. At the outset of a program, a test and evaluation master plan is developed, which describes T&E activities over a program’s life cycle—including developmental test and evaluation (DT&E), operational test and evaluation (OT&E), and potentially live-fire test and evaluation (LFT&E) at different phases—and identifies evaluation criteria for the testers. This approach is not well suited for ML/DL, which requires a more agile, iterative development and testing approach. With ML/DL systems, development is never really finished, so neither is testing. Further, ML/DL system performance is difficult to characterize and bind, and the brittleness of such systems means they will require regular system updates and testing. Exhaustive up-front testing does not make sense for these types of non-determinative systems. Therefore, the Defense Department must embrace the commercial best practice of Development, Security, and Operations (DevSecOps), a collection of processes, principles, and technologies that enables an integrated and automated approach to development and testing.9

**TEVV fails – it relies on outdated methods which don’t apply to AI – plan doesn’t change the Methodology of TEVV**

**Scharre, 2016 -- Vice President and Director of Studies at CNAS** [Paul, Feb 2016, Center for New American Security, “Autonomous weapons and operational risk”, https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS\_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515&focal=none, Acc 6/21/22, M. A.]

System failures: System failures occur not from the breakdown of any one given part, but from unanticipated interactions between elements of a system. Verifying all possible combinations of the internal workings of the system becomes increasingly difficult as the system’s complexity increases. o A recent report on autonomy by the U.S. Air Force Office of the Chief Scientist highlighted the need for new techniques for the verification and validation of autonomous software as a “critical” issue for the Air Force. “Traditional methods … fail to address the complexities associated with autonomy software … There are simply too many possible states and combination of states to be able to exhaustively test each one.”14 11 Emergent behavior could come from individual systems or from groups or swarms of simpler systems coordinating their actions together, similar to ants, termites, or bees. For more on military applications of swarming, see Paul Scharre, “Robotics on the Battlefield – Part II: The Coming Swarm,” Center for a New American Security, October 2014,

### --Extend - Fragmentation

**No solvency – the plan doesn’t account for fragmented military AI programs between different services.**

**Tarraf, Shelton, Parker 2019 - Senior Information Scientist, Senior Engineer, and physical Scientist at the RAND Corporation** [ Danielle C., William, Edward, et al RAND Cooperation, “The Department of Defense Posture for Artificial Intelligence”, <file:///Users/MiraAgarwal/Downloads/RAND_RR4229%20(2).pdf>, Acc 6/18/22, M.A.]

Within the Services The services have also all had significant activity in AI over the past year. The Army stood up an AI Task Force under the newly established AFC and established an Army AI Hub consisting of a consortium of industry, government, and academia partners based at Carnegie Mellon University (CMU).10 The Air Force stood up an AI CFT 10 See, for example, the reporting on the establishment of the Army AI hub at CMU (Matthew Nagel, “Army AI Task Force Selects Carnegie Mellon as New Hub,” Carnegie Mellon DoD Posture for Artificial Intelligence 49 and launched the Massachusetts Institute of Technology–Air Force AI Accelerator.11 The Navy and Marines stood up AI task forces. The Department of the Navy is also currently in the midst of a significant reorganization that will likely affect its posture for AI.12 The Marines, as part of the Department of the Navy, are also leveraging the Navy’s efforts At the time of our interviews, neither the Navy nor the Marines had AI-specific initiatives or partnerships with universities. The Army, Air Force, and Marines developed AI strategy annexes per the OSD directive. Of these, only the Army and Air Force strategies are publicly available.13 We identified the following set of impediments and friction points in organization, strategy, and resourcing at the level of the individual services. The service AI annexes lack baselines and metrics. All the (public) service annexes lack such metrics. The Army AI Strategy Annex presents an overarching strategy for the Army,14 decomposed in terms of ends (goals of the strategy), ways (underlying methods and an initial set of projects), and means (the manner in which the strategy will be implemented).15 The strategy mentions a forthcoming integration plan with more detail on organization, methods, and implementation, but the present strategy does not present metrics or quantifiable measures to assess progress toward the ends. Likewise, the Air Force AI University, blog post, December 4, 2018). 11 See, for instance, the reporting on the establishment of the MIT-Air Force AI Accelerator hub (Rob Matheson, “MIT and U.S. Air Force Sign Agreement to Launch AI Accelerator,” MIT News, blog post, May 20, 2019). 12 We heard of plans to set up a Navy AI Task Force but were unable to confirm its existence at the time of the report. 13 Under Secretary of the Army, “Army Artificial Intelligence Strategy Annex Submission,” memorandum for Chief Information Office, Office of the Secretary of Defense, Washington, D.C.: U.S. Department of Defense, 2019; U.S. Department of the Air Force, The United States Air Force Artificial Intelligence Annex to the Department of Defense Artificial Intelligence Strategy, Washington, D.C., 2019. 14 Although not exclusively focused on operational AI and mission-support AI, the Army AI strategy seems to emphasize those over enterprise AI. 15 Under Secretary of the Army, 2019; U.S. Department of the Air Force, 2019. 50

### --Extend – Safety Impossible

**Safety cannot solve military AI malfunctions – accelerated speed and loss of human control are intentional and irreversible**

**Horowitz and Scharre, 2021 - Director of the Emerging Capabilities Policy Office in the Office of the Under Secretary of Defense for Policy and  Vice President and Director of Studies at CNAS** [Michael, Paul, January 12 2021, “AI and International Stability: Risks and Confidence- Building Measures”, <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>, Acc 6/18/22. M.A.]

Even if militaries successfully manage safety and security concerns and field AI systems that are robust and secure, properly functioning AI systems could create challenges for international stability. For example, both Chinese and American scholars have hypothesized that the introduction of AI and autonomous systems in combat operations could accelerate the tempo of warfare beyond the pace of human control. Chinese scholars have referred to this concept as a battlefield “singularity,”[7](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn7) while some Americans have coined the term “hyperwar” to refer to a similar idea.[8](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn8) If warfare evolves to a point where the pace of combat outpaces humans’ ability to keep up, and therefore control over military operations must be handed to machines, it would pose significant risks for international stability, even if the delegation decision seems necessary due to competitive pressure. Humans might lose control over managing escalation, and war termination could be significantly complicated if machines fight at a pace that is faster than humans can respond. In addition, delegation of escalation control to machines could mean that minor tactical missteps or accidents that are part and parcel of military operations in the chaos and fog of war, including fratricide, civilian casualties, and poor military judgment, could spiral out of control and reach catastrophic proportions before humans have time to intervene.

**TEVV cannot eliminate the risk of Accidents – systems are too complex and test conditions cannot cover every environment.**

**Scharre, 2016 -- Vice President and Director of Studies at CNAS** [Paul, Feb 2016, Center for New American Security, “Autonomous weapons and operational risk”, https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/CNAS\_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515&focal=none, Acc 6/21/22, M. A.]

From an operational standpoint, autonomous weapons pose a novel risk of mass fratricide, with large numbers of weapons turning on friendly forces. This could be because of hacking, enemy behavioral manipulation, unexpected interactions with the environment, or simple malfunctions or software errors. Moreover, as the complexity of the system increases, it becomes increasingly difficult to verify the system’s behavior under all possible conditions; the number of potential interactions within the system and with its environment is simply too large. While these risks can be mitigated to some extent through better system design, software verification and validation, test and evaluation, and user training, these risks cannot be eliminated entirely. Complex tightly coupled systems are inherently vulnerable to “normal accidents.” The risk of accidents can be reduced, but never can be entirely eliminated.

### --Extend – JAIC Fails

**No solvency – the DOD will not enhance JAIC authority to solve the case – empirically, they have fragmented acquisitions.**

**Tarraf, Shelton, Parker 2019 - Senior Information Scientist, Senior Engineer, and physical Scientist at the RAND Corporation** [ Danielle C., William, Edward, et al RAND Cooperation, “The Department of Defense Posture for Artificial Intelligence”, <file:///Users/MiraAgarwal/Downloads/RAND_RR4229%20(2).pdf>, Acc 6/18/22, M.A.]

What DoD needs to do now is continue on that path by providing the requisite high-level support, visibility, and authorities (including directive and budget authorities) to enable the JAIC to enact change. Doing so would ensure that the JAIC has a chance of succeeding at its mandate of scaling AI and its impact across DoD. It would also ensure that DoD’s intent, messaging, and actions are all consistent. Having said that, we recognize that this option runs counter to DoD history and precedents, particularly because of the recent reform leading to the dissolution of the USD for Acquisition, Technology, and Logistics, and subsequent creation of the USD(R&E) and USD(A&S) 3 The DMAG is the primary civilian-military management forum that supports the Secretary of Defense and addresses top DoD issues that have resource, management, and broad strategic and/or policy implications. The DMAG’s primary mission is to produce advice for the DSD in a collaborative environment and to ensure that the DMAG execution aligns with the Secretary of Defense’s priorities and the planning and programming schedule. The DMAG is cochaired by the DSD and Vice Chairman of the Joint Chiefs of Staff, with secretaries of the military departments, chiefs of the military services, and DoD principal staff assistants holding standing invitations. See U.S. Department of Defense, Chief Management Officer, “Deputy’s Management Action Group (DMAG),” webpage, undated. 68 The Department of Defense Posture for Artificial Intelligence that took effect in February 2018. By enacting this reform, Congress intentionally weakened the directive authorities that OSD principals had over the services, and devolved significant procurement and acquisition authorities back to the services. We also recognize that it might not be entirely appropriate to compare DoD with a large company but rather to a large conglomerate because of the historical role and independence of the services.

#### The Plan will fail - lack of JAIC authority makes budgeting impossible

Tarraf, Shelton, Parker 2019 - Senior Information Scientist, Senior Engineer, and physical Scientist at the RAND Corporation [ Danielle C., William, Edward, et al RAND Cooperation, “The Department of Defense Posture for Artificial Intelligence”, <file:///Users/MiraAgarwal/Downloads/RAND_RR4229%20(2).pdf>, Acc 6/18/22, M.A.]

The JAIC lacks the authorities to carry out its present role. At its core, the JAIC’s overarching mission can be distilled to this: Scale AI and its impact across DoD. This mission and its present scope—as defined by the summary of the DoD AI strategy and the memo establishing the JAIC—are extensive, while the JAIC’s current authorities are limited. In particular, the JAIC is expected to synchronize DoD AI activities and coordinate AI initiatives totaling more than $15 million annually. It is unclear whether the JAIC has any mechanisms for enforcing these directives, because it does not have the authorities to direct investments or to halt programs or activities that are deemed to be misaligned with DoD’s strategy (a fact we learned through multiple DoD interviews). In short, the JAIC does not have directive or budget authorities, and that critically limits its ability to synchronize and coordinate DoD-wide AI activities to enact change. Currently, it can catalogue these activities, but it is unclear how doing so would help scale AI across DoD. Of course, that assumes that what constitutes an AI activity is known. However, it is not currently clear how the determination of what constitutes an AI initiative or activity is made, by whom, and whether that determination is consistent across DoD.8

**JAIC fails – lack of long term funding certainty undermines industry support**

**Tarraf, Shelton, Parker 2019 - Senior Information Scientist, Senior Engineer, and physical Scientist at the RAND Corporation** [ Danielle C., William, Edward, et al RAND Cooperation, “The Department of Defense Posture for Artificial Intelligence”, <file:///Users/MiraAgarwal/Downloads/RAND_RR4229%20(2).pdf>, Acc 6/18/22, M.A.]

The lack of longer-term budget commitments might hinder the JAIC’s success. This observation is not just about the amount of funding for the JAIC—for which we have no basis to judge at present— but also the horizon, certainty (or lack thereof), and general trends of funding commitments. Our insights gleaned from industry indicate that a sizable, long-term funding commitment, generally ramping up to accompany the five-year strategic road map, is critical to ensuring success in organizational transformations to enable scaling of AI (see section “Industry: Organization” in Appendix C). Based on our interactions with the JAIC, we were unable to determine whether the JAIC is able to submit budget requests through the programming, planning, budgeting and execution (PPBE) system as an independent entity, allowing it to request funds for the Future Years Defense Plan (FYDP) and also allowing high-level leadership to demonstrate support for the JAIC’s mission by prioritizing these budget requests.

### --Extend – Training Data

**Alternate Causality - Limited DOD data prevent testing and verification of AI**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

Testing ML/DL requires large, representative data sets. While technological advances in “one shot” and reinforcement learning may ultimately enable the Pentagon to test ML/DL without a lot of data or provide alternative approaches to handle out-of-distribution situations, for the next five to 10 years, the Defense Department will likely rely on supervised learning systems, and testing ML/DL systems will likely require large sets of labeled, representative data. The United States needs a whole-of-government data strategy that allows for data collection, cleaning, curation, and sharing across agencies, especially between DOD and the IC. Currently, the Defense Department lacks sufficient available data that mimics the conflict condition in which these systems may operate in the future. This will limit its ability to test system performance against realistic conditions. It will also hamstring efforts to identify edge cases and develop fail-safe mechanisms to prevent catastrophic outcomes. The Pentagon lacks the ability to effectively collect, manage, store, and share testing data across the enterprise, which would enable this approach to scale. Finally, DOD leadership will need approaches to continuously test the quality of the data itself, as testing data could be compromised or revealed unintentionally or intentionally by adversaries. ML/DL will be integrated into a system of systems. ML/DL will be integrated into a range of DOD software and hardware systems, so it is imperative that developers, testers, and policymakers take a systems architecture view when building and evaluating these systems. The Defense Department cannot simply test all components separately and assume that the system as a whole will work as intended. The accuracy and precision of ML/DL systems is typically a composite effect that arises from a combination of the behaviors of different components, such as the training data, the learning program, and even the learning framework. These components are then embedded in larger systems, so interactions with the physical, computational, and human components of the system will ultimately affect system performance. Often, failures come from unexpected interactions or relationships between systems, rather than the behavior of any individual element. These dynamics make the system increasingly vulnerable to malfunction and cyber-attacks. An adversary could attack any number of vulnerable entry points within the hardware or software that could, in turn, compromise the entire system.11 The Defense Department needs to greatly advance its ability to conduct integrated systems testing that takes into account the interactions with and between systems, testing both machine-machine and human-machine interactions. It should also prioritize testing for how failure in a given subsystem could impact the performance of the system as a whole.

#### Effective AI requires enormous amounts of data inputs, which are less available and more protected in the military sphere.

Goldfarb and Lindsay, 2022 – Chair in AI and health care at the Univ of Toronto, Prof of Cybersecurity at the Georgia Tech [Avi, Jon, 2/25/22, <https://doi.org/10.1162/isec_a_00425>, “Prediction and Judgement: Why Artificial Intelligence Increases the Importance of Humans in War” 6/18/22, LND]

Michael Horowitz describes AI as “the ultimate enabler” for automating decision-making tasks in everything from public administration and commercial business to strategic intelligence and military combat.[5](javascript:;) In 2018, the Department of Defense observed that “AI is poised to transform every industry, and is expected to impact every corner of the Department, spanning operations, training, sustainment, force protection, recruiting, healthcare, and many others.”[6](javascript:;) We would be surprised, however, if AI transformed all these activities to the same degree for all actors who use it. One of the key insights from the literature on the economics of technology is that the complements to a new technology determine its impact.[7](javascript:;) AI, from this perspective, is not a simple substitute for human decision-making. Rapid advances in machine learning have improved statistical prediction, but prediction is only one aspect of decision-making. Two other important elements of decision-making—data and judgment—represent the complements to prediction. Just as cheaper bread expands the market for butter, advances in AI that reduce the costs of prediction are making its complements more valuable. AI prediction models require data, and accurate prediction requires more and better data. Quality data provide plentiful and relevant information without systemic bias. Data-driven machine prediction can efficiently fill in information needed to optimize a given utility function, but the specification of the utility function ultimately relies on human judgment about what exactly should be maximized or minimized. Judgment determines what kinds of patterns and outcomes are meaningful and what is at stake, for whom, and in which contexts. Clear judgments are well specified in advance and agreed upon by relevant stakeholders. When quality data are available and an organization can articulate clear judgments, then AI can improve decision-making. We argue that if AI makes prediction cheaper for military organizations, then data and judgment will become both more valuable and more contested. This argument has two important strategic implications. First, the conditions that have made AI successful in the commercial world—quality data and clear judgment—may not be present, or present to the same degree, for all military tasks. In military terms, judgment encompasses command intentions, rules of engagement, administrative management, and moral leadership. These functions cannot be automated with narrow AI technology. Increasing reliance on AI, therefore, will make human beings even more vital for military power, not less. Second, the importance of data and judgment creates incentives for strategic competitors to improve, protect, and interfere with information systems and command institutions. As a result, conflicts over information will become more salient, and organizational coordination will become more complex. In contrast with assumptions about rapid robot wars and decisive shifts in military advantage, we expect AI-enabled conflict to be characterized by environmental uncertainty, organizational friction, and political controversy. The contestation of AI complements, therefore, is likely to unfold differently than the imagined wars of AI substitutes.[8](javascript:;)

### --Extend - Workforce

**No solvency – the US lacks the AI workforce for effective TEVV**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

9. Accelerate recruitment and training of ML/DL and TEVV talent. Recruiting and retaining diverse, interdisciplinary teams is an essential prerequisite to advancing TEVV for ML/DL systems. DoD needs those with a fundamental academic grounding in test and evaluation, as well as the systems engineers, computer scientists, and ML/DL experts that understand the technology itself. It also needs statisticians, data scientists, and applied mathematicians who can perform mathematical testing. Finally, it needs experts who understand human-machine interaction, such as psychologists and ethicists. In addition to these subject matter experts, it also needs operators, requirements writers, acquisition professionals, and lawyers who have a basic degree of technological literacy and understand why this technology matters.

**No Solvency – The DOD lacks an AI workforce due to slow hiring and low pay**

**Tarraf et. al 2019 - Senior Information Scientist at the RAND Corporation** [Danielle with William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus, Justin Grana, Alexis Levedahl, Jasmin Léveillé, Jared Mondschein, James Ryseff, “The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations”, LMSi]

DoD struggles to grow and cultivate AI talent. Our interviews suggest a mixed appreciation for what technical AI talent consists of and which AI talent is needed. Several entities we interviewed, such as the service labs, had a clear sense of AI talent needs, but the majority were still in the beginning stages of such considerations and were more likely to emphasize contracting out for technical talent. Moreover, for those that were clear on AI talent needs, it was a challenge to define the exact knowledge, skills, and abilities they perceived that were required. Ultimately, the AI talent needs of DoD (type,38 quantity, and mix) will depend on the broader strategy pursued for scaling AI, and the extent to which scaling AI will rely on the development of products in-house as opposed to through contracting and outsourcing. The skill sets needed for development of products in-house are significantly different from those needed for contracting and outsourcing, though all AI talent (technical or managerial) is difficult to access in the present market. Nonetheless, the consensus is that DoD faces stiff competition for AI skills and expertise, as evidenced by our interviews across academia, industry, and DoD.39 Many of our DoD interviewees discussed the challenges related to attracting and recruiting technical talent more generally, and expressed the belief that AI talent would be no different. In that spirit, we point to a recent RAND study on career paths for data scientists within the Defense Intelligence Agency.40 Interviews across DoD cited intense competition with the private sector, the limited ability to compete on salary, and long hiring processes. At the same ML developers (see section “Industry: Talent” in Appendix C), that approach will not lead to the development of ML experts. 38 Our industry interviews highlighted four types of AI talent: experts, ML developers, application developers, and project or program managers (see section “Industry: Talent” in Appendix C). 39 Reasons for such stiff competition include salaries and inability to hire at competitive speed.

### --Extend – No Definition

**The DOD cannot implement an AI policy because there is no precise definition of AI to guide programs.**

**Tarraf, Shelton, Parker 2019 - Senior Information Scientist, Senior Engineer, and physical Scientist at the RAND Corporation** [ Danielle C., William, Edward, et al RAND Cooperation, “The Department of Defense Posture for Artificial Intelligence”, <file:///Users/MiraAgarwal/Downloads/RAND_RR4229%20(2).pdf>, Acc 6/18/22, M.A.]

Currently, it can catalogue these activities, but it is unclear how doing so would help scale AI across DoD. Of course, that assumes that what constitutes an AI activity is known. However, it is not currently clear how the determination of what constitutes an AI initiative or activity is made, by whom, and whether that determination is consistent across DoD.8 The JAIC lacks a five-year strategic road map, and a precise objective allowing it to formulate one. Our industry interviews (see section “Industry: Organization in Appendix C) and relevant literature highlight the need for five-year strategic road maps to execute organizational transformation,9 particularly a transformation of the magnitude envisioned in the DoD AI strategy and that the JAIC has been tasked with executing. In that context, our industry interviews also emphasized the need for an objective articulated in precise-enough terms to enable the formulation of such a strategic road map (see section “Industry: Organization” in Appendix C). DoD experience with technology also highlights the importance of clearly defined, measurable goals in 8 We touched upon this point earlier in Chapter Three while discussing the definition of AI. Although it is not clear that enforcing a DoD-wide definition of AI is either feasible or helpful, the question of how DoD identifies and tracks AI activities or programs remains an important open question. 9 John M. Bryson, Lauren Hamilton Edwards, and David M. Van Slyke, “Getting Strategic About Strategic Planning Research,” Public Management Review, Vol. 20, No. 3, 2018. 48 The Department of Defense Posture for Artificial Intelligence enhancing success (see section “Adoption and Scaling of Unmanned Aircraft Systems” in Appendix D). The JAIC’s mission, which we have distilled to scale AI and its impact across DoD, is too vague to serve as a five-year objective for the purpose of this road map. The JAIC needs a refined objective that is precise, ambitious, and potentially feasible in the time frame, and that can serve to guide the development of an agile, strategic road map to include shorter-term (one-year) goals and metrics to assess progress along these goals. The existence of a five-year strategic road map would also help focus the selection of NMIs and justify their relevance to the overall objective (see “Organization: At OSD Level” in Appendix B).

### --Extend – Industry Cooperation

**Alternate causality – America’s computer industry refuses to cooperate with the military on AI**

**Scharre, 2019 - Vice President and Director of Studies at CNAS** [Paul, May-June, “Killer Apps: The Real Dangers of an AI Arms Race,” https://omnilogos.com/killer-apps-real-dangers-of-ai-arms-race/6/18/22 MD]

Equally alarming for U.S. policymakers is the sharp divide between Washington and Silicon Valley over the military use of AI. Employees at Google and Microsoft have objected to their companies' contracts with the Pentagon, leading Google to discontinue work on a project using AI to analyze video footage. China's authoritarian regime doesn't permit this kind of open dissent. Its model of "military-civil fusion" means that Chinese technology innovations will translate more easily into military gains. Even if the United States keeps the lead in AI, it could lose its military advantage. The logical response to the threat of another country winning the AI race is to double down on one's own investments in AI. The problem is that AI technology poses risks not just to those who lose the race but also to those who win it.

### Military Readiness Links

**Strict testing and reliability for AI hurts military deployment**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

DOD will first need to establish a testing framework that provides guidance on the degree of acceptable risk and limits for a given ML/DL use case based on a potential range of outcomes and errors. For example, if a potential outcome has lethal consequences, the acceptable risk is likely to be extremely low, whereas if the outcome has no clear negative consequences, the acceptable risk will almost certainly be higher. The risk of fielding these systems will also need to be weighed against the risk associated with not adopting the system. For example, a 5 percent error rate may be palatable if the existing system has a 10 percent error rate. These risk and error rates will also need to incorporate the potential for adversary attacks or interactions with adversary systems. For example, an error that happens .001 percent of the time naturally, but which an adversary is able to consistently exploit, could create significant challenges for the Pentagon. Further, policymakers must acknowledge that with technology, there might be less margin for error than with humans, and less clarity about who is accountable for such errors. For example, the United States may determine that as a society, we are not willing to accept a scenario in which an algorithmic error in an autonomous vehicle causes a loss of life even if it saves thousands of lives overall. Ultimately, these technologies will never be perfect, and testing to a near-perfect standard will inhibit DOD’s ability to field these systems at all. Therefore, it needs a dedicated process to develop policies to determine how much risk it is willing to accept in a given case, weighing operational need and potential consequences against DOD ethics, principles, and policies. DOD will need to translate this testing and safety framework into functional, specific requirements language. For example, the JAIC could put out a request for proposal saying it needs a DL that can identify a target from X range, in this season, in these weather conditions.

**Current military efforts are rapidly scaling up AI use due to flexibility – The plan distracts from this by committing to a particular initiative**

**Reilly 2022 – Emerging tech reporter for Inside Defense** [Briana, June 8, “Hicks: One Year in, AI Adoption Initiative ‘Proving Its Worth’”, https://www-proquest-com.proxy.lib.umich.edu/docview/2676524362?pq-origsite=primo]

One year into the Pentagon's artificial intelligence and data accelerator initiative, Deputy Defense Secretary Kathleen Hicks says the effort is already "proving its worth." While it's unclear what will happen to the three-year push after fiscal year 2024, Hicks told an online audience during DOD's Digital and AI Symposium today that in the interim, ADA has already allowed officials to identify common problems faced by the combatant commands in integrating and scaling AI capabilities. In that same vein, Hicks added, officials have also been able to identify "common solution approaches" to the initiative launched last summer as part of the military's approach to enabling Joint All-Domain Command and Control. "Because we're at the enterprise level at ADA, that's kind of the beauty of the federated approach," she continued. "We have this centralized repository of knowledge and expertise and data and tools and contract vehicles and folks who understand how to use contract vehicles for this purpose. And then these problem sets can come in and we can tailor, if you will, more easily and get solutions out faster. I think that's what we've seen to date." Through ADA, DOD has been sending technical AI expert teams to the COCOMs as one of its initial steps, the former acting Pentagon chief information officer, Kelly Fletcher, previously said. Hicks noted today that COVID-19 has complicated that recently. Regardless, she said that the department has witnessed "a lot of natural-use cases [and] questions that combatant commanders really want help answering, and the ability to apply answers through ADA." ADA is poised to represent "a flagship achievement" for the chief data and artificial intelligence officer, in the words of DOD CIO John Sherman, who spoke separately during the symposium today. The newly stood-up CDAO, which consolidates the military's data and AI efforts in an attempt to provide better alignment for the Pentagon, will be able to "surface up certain trends, certain enterprise-level activities and needs that will transcend the commands" through ADA, Sherman said. As the department prepares to move beyond ADA in two years and explore what that "next natural evolution" might be, Hicks stressed the importance of being "very unafraid to shift approaches, as the stand-up of the CDAO itself shows, and make sure we are ahead of the curve, not chasing a curve by being committed to either particular initiatives and/or to organizational constructs." "I think anything that would follow it, whether it's called ADA or something else, naturally will build on what we do here," she added.

### Spending Links

**Reforming DOD TEVV process will require substantial investments and attaching new funding mandates to each AI project.**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

An agile approach of iterative testing, updates, and releases will place significant burdens on TEVV and require infrastructure and research investments, as well as incentivizing program managers to see testing as an integral part of the development process rather than a barrier. Program managers should be responsible and rewarded for delivering a well-functioning product, not just staying on budget and schedule. Current TEVV methods and infrastructure aren’t well suited for ML/DL and may require new funding approaches. Adapting the TEVV enterprise for ML/DL will require targeted investment in developing new testing methods and adapting current testing infrastructure to support DevSecOps and iterative testing. The Defense Department needs new approaches, such as automated testing and digital twinning,18 as well as new testing infrastructure, including test beds, test ranges, and advanced modeling and simulation (M&S). DOD also needs computing support, cloud-based resources, data capture for continuous development, and generation and use of synthetic data, particularly for DL applications.19 Finally, it needs tools for traceability that capture key information about the systems development and deployment to inform follow-on development, testing, and use. The JAIC has adopted commercial best practices for AI DevSecOps. Its Joint Common Foundation (JCF)—an infrastructure environment designed specifically for training, testing, and transitioning AI technologies, which is intended for use by all the services—is an important down payment on these efforts that will make it easier to secure and rapidly test and authorize AI capabilities.20 The JAIC should be given the resources and top-cover it needs to scale this effort. The Pentagon should build on the JCF and other efforts to promote a secure, cloud-based DevSecOps ecosystem that facilitates the rapid commercial development and iterative testing of ML/ DL and the proliferation of testing tools, data, and standards across OSD and the services. The Defense Department also needs to increase resources, bandwidth, and personnel dedicated to adversarial testing. It can and does use Federally Funded Research and Development Centers (FFRDCs), but there is concern among some experts that it is too heavily reliant on just one—MITRE—for adversarial testing. DOD needs to invest in creating a catalogue of adversarial testing tools and proliferate these capabilities across the service labs and FFRDCs that support testing. Finally, DoD needs to work more closely with the intelligence community to simulate realistic threats. Finally, new approaches to TEVV for ML/DL will require new funding approaches. DoD, in coordination with Congress, should consider new approaches that incorporate T&E funding into the cost of development, given that TEVV must be integrated into an iterative development process. DoD and Congress should also consider establishing a new appropriations category that allows AI/ML to be funded as a single budget item, with no separation between RDT&E, production, and sustainment, as recommended by the Defense Innovation Board Software Acquisition and Practices Study.21

**Plan will require Congress to remove spending caps and establish new budget categories**

**Flournoy, Haines et al 2020 - former Under Secretary of Defense for Policy and Director of National Intelligence** [Michèle and Avril, Gabrielle Cheftiz Special Assistant to the Under Secretary of Defense for Policy October, “Building Trust through Testing Adapting DOD’s Test & Evaluation, Validation & Verification (TEVV) Enterprise for Machine Learning Systems, including Deep Learning Systems” https://cset. georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf Acc 6/23/22 JZ]

6. Increase and integrate spending for T&E research and infrastructure. Advancing TEVV for ML/DL will require a substantial investment in both research and infrastructure. TRMC should lead on assessing current gaps in infrastructure and be given increased funds to invest in service and DoD T&E live, virtual, and constructive (LVC) test ranges, test beds, and modeling and simulation for testing adaptive systems. DoD should significantly increase investment in modeling, digital twins, and simulation, working with the private sector—particularly commercial autonomous vehicle companies—to implement industry best practices. These technologies can be used to develop representative testing environments and conduct edge testing to determine a system’s operational envelope. This investment is also key to reaching the goal of automatic, repeatable testing, which is critical to DevSecOps, and creating synthetic data that can help offset a lack of usable, operational data. DoD could invest in test beds to be hosted at FFRDCs and university-affiliated research centers, which attract top talent and work with DoD regularly. TRMC should also help scale the Navy’s automatic test and retest program, which uses cloud-based digital twins to provide near real-time feedback and automatic testing of thousands of simulated environments.31 To do so, we support the National Security Commission on AI’s recommendation that Congress should raise the authorized cap for laboratory infrastructure investments, currently set at $6 million, in order to provide laboratories with the ability to invest in equipment and testbed infrastructure necessary for robust AI research, prototyping, and testing.32 Finally, the Department should consider new approaches to fund AI/ML TEVV. For example, DoD could require that TEVV cost is factored into development, rather than having as a separate T&E item. Congress could also consider a new type of funding authority that bridges the gap between AI S&T and T&E, allowing for both development and testing of new technology. DoD does not yet have well-established methods of testing for ML/DL, and will therefore be developing the capability and the ability to test it in parallel. This will require S&T dollars for research on new T&E approaches. Congress has already authorized a similar model for cyber, in which funds are authorized for creating, testing, fielding, and operations.33 The Department, working with Con25 gress, should explore the potential of replicating this model for AI development and testing, consistent with the Defense Innovation Board Software Acquisition and Practices study recommendation for a single budget item for AI/ML.34

### Political Capital Links

**Reforming the testing and evaluation system will cost political capital because it regulates defense industries**

**Gilli, 2020 - Senior Researcher at the NATO Defense College** [Andrea, NDC Research Paper No.15 – December ““NATO-Mation”: Strategies for Leading in the Age of Artificial Intelligence” https://www.ndc.nato.int/news/news.php?icode=1514 Acc 4/21/22 TA]

Defence procurement is particularly complex, given the range of strict requirements and specifications with which contractors have to comply.251 This is why traditional weapons manufacturers generally have an advantage when it comes to approaching defence buyers.252 AI, however, will entail a paradigm shift because the capabilities will continuously evolve. For AI-enabled autonomous systems in particular, as mentioned above, decision-making is “non-deterministic” and depends on the “dynamic environment” in which the system operates.253 This means that “traditional development and procurement approaches, based on full-path regression, are unfit”.254 This means that R&D also continues in the product deployment phase, as data and algorithms keep providing information and feedback that have to be integrated in order to achieve initial operational capabilities. Current “waterfall” procurement paradigms are set up so that engineers test prototypes according to defined specifications, and then subsequently move to production.255 With machine learning, the specifications will keep evolving as algorithms are fed new data.256 By extension, this means that testing cannot be treated as a singular phase prior to production and development. For NATO Allies, a more iterative procurement paradigm that does not depend on the sequential “waterfall” entails unique challenges. The integration of enterprise AI requires specific technical, legal, and organizational capabilities. This is linked to human-capital challenges in requirements, procurement and operations communities, as organizations leveraging AI “need their own people who know how to structure the problem, handle the data, and stay aware of evolving opportunities”.257 The development of AI-centred major weapon systems similarly calls for reform, or at least a remarkable adaptation, of defence procurement.258 Significant political, organizational and human capital will have to be invested. This is an additional reason why a centre such as the proposed A3IC could support the transition, through best practices, lessons learned and similar initiatives.

### AIA Counterplan – 1NC

#### Text – The United States should enter into an Autonomous Incidents Agreement with the People's Republic of China.

**An Autonomous Incidents Agreement minimizes unintended escalation – precedent shows rules of the road can avoid military accidents**

**Horowitz and Scharre 2021 - Director of the Emerging Capabilities Policy Office in the Office of the Under Secretary of Defense for Policy and  Vice President and Director of Studies at CNAS** [Michael, Paul, January 12 2021, “AI and International Stability: Risks and Confidence- Building Measures”, <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>, Acc 6/18/22. M.A.]

There are inherent risks when autonomous systems with any level of decision-making interact with adversary forces in contested areas. Given the brittleness of algorithms, the deployment of autonomous systems in a crisis situation could increase the risk of accidents and miscalculation. AI-related CBMs could build on Cold War agreements to reduce the risk of accidental escalation, with some modification to account for the new challenges AI-enabled autonomous systems present. States have long used established “rules of the road” to govern the interaction of military forces operating with a high degree of autonomy, such as at naval vessels at sea, and there may be similar value in such a CBM for interactions with AI-enabled autonomous systems. The 1972 Incidents at Sea Agreement and older “rules of the road” such as maritime prize law provide useful historical examples for how nations have managed analogous challenges in the past. Building on these historical examples, states could adopt a modern-day “international autonomous incidents agreement” that focuses on military applications of autonomous systems, especially in the air and maritime environments. Such an agreement could help reduce risks from accidental escalation by autonomous systems, as well as reduce ambiguity about the extent of human intention behind the behavior of autonomous systems. In addition to the Incidents at Sea Agreement, maritime prize law is another useful historical analogy for how states might craft a rule set for autonomous systems’ interactions. Prize law, which first began in the 12th century and evolved more fully among European states in the 15th to 19th centuries, regulated how ships interacted during wartime. Because both warships and privateers, as a practical matter, operated with a high degree of autonomy while at sea, prize law consisted of a set of rules governing acceptable wartime behavior. Rules covered which ships could be attacked, ships’ markings for identification, the use of force, seizure of cargo, and providing for the safety of ships’ crews.[61](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn61) Nations face an analogous challenge with autonomous systems as they become increasingly integrated into military forces. Autonomous systems will be operating on their own for some period of time, potentially interacting with assets from other nations, including competitors, and there could be value in establishing internationally agreed upon “rules of the road” for how such systems should interact. The goal of such an agreement, which would not have to be as formal as the Incidents at Sea Agreement, would be to increase predictability and reduce ambiguity about the behavior of autonomous systems. Such an agreement could be legally binding but would not necessarily need to be in order to be useful. It would likely need to be codified in an agreement (or set of agreements), however, so that expectations are clear by all parties. An ideal set of rules would be self-enforcing, such that it is against one’s own interests to violate them. Examples of rules of this kind in warfare include prohibitions against perfidy[62](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn62) and giving “no quarter,”[63](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn63) violating either of which incentivizes the enemy to engage in counterproductive behavior, such as refusing to recognize surrender or fighting to the bitter end rather than surrendering. An autonomous incidents agreement could also include provisions for information-sharing about potential deployments of autonomous systems in disputed areas and mechanisms for consultation at the military-to-military level to resolve questions that arise (including potentially a hotline to respond to incidents in real time).

**An Autonomous Incidents Agreement solves escalation by monitoring compliance – China will be willing to cooperate.**

**Horowitz and Scharre 2021 - Director of the Emerging Capabilities Policy Office in the Office of the Under Secretary of Defense for Policy and  Vice President and Director of Studies at CNAS** [Michael, Paul, January 12 2021, “AI and International Stability: Risks and Confidence- Building Measures”, <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>, Acc 6/18/22. M.A.]

One benefit to an international rule set that governs the behavior of autonomous systems, particularly in peacetime or pre-conflict settings, is that the outward behavior of the system would be observable, even if its code is not. Other nations could see how another country’s autonomous air, ground, or maritime drone behaves and whether it is complying with the rules, depending on how the rules are written. One benefit to an international rule set that governs the behavior of autonomous systems, particularly in peacetime or pre-conflict settings, is that the outward behavior of the system would be observable, even if its code is not. Given the perceived success of the Incidents at Sea Agreement in decreasing the risk of accidental and inadvertent escalation between the United States and the Soviet Union, an equivalent agreement in the AI space might have potential to do the same for a new generation. The efficacy of any agreement would depend on the details, both in the agreement itself and in states’ execution. For example, the United States and China have signed multiple CBM agreements involving air and maritime deconfliction of military forces, including the 1998 U.S.-China Military Maritime Consultative Agreement and the 2014 Memorandum of Understanding Regarding the Rules of Behavior for Safety of Air and Maritime Encounters, yet U.S.-China air and naval incidents have continued.[64](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn64) However, the existence of these prior agreements themselves may be a positive sign about the potential for U.S.-China cooperation on preventing accidents and could be a building block for further collaboration. Moreover, in a February 2020 article, Senior Colonel Zhou Bo in China’s People’s Liberation Army (PLA) advocated for CBMs between the United States and China, including on military AI, drawing on the example of the 1972 Incidents at Sea Agreement.[65](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn65) Interest in at least some quarters in the Chinese military suggests that cooperation may be possible even in the midst of competition, especially if the PLA is willing to reciprocate American transparency.[66](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn66)

**Military to military dialogue with China prevents the worst AI accidents from escalating**

**Horowitz and Kahn, 2021 – Senior and Research Fellows at the Council on Foreign Relations** [Michael and Lauren, The Washington Quarterly 3-19-2021 “Leading in Artificial Intelligence through Confidence Building Measures”  [https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/0163660X.2021.2018794 acc on 6-21-2022](%20https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/0163660X.2021.2018794%20acc%20on%206-21-2022%20) TM]

Dialogue on AI Safety and Strategic Stability One type of CBM the United States could promote is dialogues, including scientific dialogues, Track-II dialogues, and military-to-military dialogues on issues surrounding military uses of AI, including strategic stability. Many Track-II dialogues are already underway between US and Chinese as well asUS and Russian participants.25 The Biden administration and Russian President Putin have already committed to an integrated Strategic Stability Dialogue designed to encompass all potential factors that could impact strategic stability, including emerging technologies.26 More dialogue with China on AI could also benefit the United States. In both cases, explaining how seriously the United States takes AI safety and its commitment to using AI in ways that enhance strategic stability could persuade some Chinese and Russian counterparts to take those issues more seriously themselves. At worst, making the case for a sensible, and safe, approach to military uses of AI will make the United States look more reasonable globally if China and Russia do not agree, given the way the content of these discussions almost inevitably becomes a part of global conversations. In that way, even if Chinese and Russian counterparts do not substantively engage, the attempt will still signal American prioritization of AI safety and enhance global US scientific credibility.

### --Extend – AIA Solves Accidents

**An Autonomous Incidents Agreement solves miscalculation and accidents – it is a confidence building measure that increases trust and transparency**

**Horowitz, 2021 - Professor of Political Science at the University of Pennsylvania** [Michael Horowitz, 1-26-2021, , Bulletin of the Atomic Scientists, "How Joe Biden can use confidence-building measures for military uses of AI” https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/00963402.2020.1860331 accessed on 6-18-2022]

Against the backdrop of continued development in military AI technologies, the Biden administration should work with allies to not just ensure US military leadership in AI, but decrease the risk that AI adoption creates risks for international stability. The Biden administration could consider what limits on AI development and use it could adopt without undermining US conventional military power. The secure second-strike capabilities of the United States in particular may help explain why some senior US military officials have expressed skepticism (Freedberg 2019) about some uses of AI in the nuclear weapons arena. Use limits could range from deploying nuclear weapons on autonomous platforms to decreasing appropriate human judgment surrounding the connectivity between early warning systems and nuclear weapons launches. Research over the last few years shows the potentially destabilizing effects some applications of AI might have on strategic stability when it comes to nuclear weapons (Horowitz, Scharre, and Velez-Green 2019) or cybersecurity (NSTC 2020). More broadly, the Biden administration will also have an opportunity to promote US leadership in AI through confidence-building measures – a tool best known from the Cold War. Confidence building measures could facilitate information-sharing and increase transparency between states developing AI (Horowitz, Kahn, and Mahoney 2020). One example might be an AI equivalent of the Cold War Incidents at Sea Agreement (US Government 1972) – which dealt with reducing the risk of maritime accidents and inadvertent escalation involving the United States and the Soviet Union – whereby countries committed to disclosing the deployment of AI-enabled platforms, so that if accidents happen (Ciocca and Kahn 2020), they are not viewed as deliberate acts that then trigger escalation. These measures offer an attractive option for initiating the process of risk mitigation as they focus on trust building rather than the sensitive root causes of conflict. It is critical not to exaggerate the likely gains from confidence-building measures, though. While shared interests exist to prevent accidental war, countries such as China and Russia are pursuing military uses of AI for their own reasons – and in part to challenge the United States. Thus, it is important to be realistic about what these measures can achieve. The Biden administration will have to make several critical decisions about the level of US military investment in AI, the scope of those investments, and the extent to which the United States will prioritize increasing international cooperation on reducing the risk of accidental or inadvertent conflicts associated with AI. These choices will be all the more difficult given that they will be occurring under a shadow of uncertainty concerning technical progress in AI. Due to this uncertainty, foreclosing the possibility of using AI in narrowly dangerous ways and initiating low-risk efforts like engaging with other countries through confidence-building measures could help ensure that the inevitable incorporation of AI by militaries around the world does not unnecessarily increase global instability.

**Rules of the Road can mitigate risks of accidents – attempts at AI Arms Control are doomed to fail empirically**

**Sharre, 2018 - director of the technology and national security program at the Center for a New American Security** [Paul, 12 September, “Ultrafast computing is critical to modern warfare. But it also ensures a lot could go very wrong, very quickly.” <https://foreignpolicy.com/2018/09/12/a-million-mistakes-a-second-future-of-war/> ST]

Attempts at arms control go back to antiquity, from the Bible’s prohibition on wanton environmental destruction in Deuteronomy to the Indian Laws of Manu that forbade barbed, poisoned, or concealed weapons. In the intervening centuries, some efforts to ban or regulate certain weapons have succeeded, such as chemical or biological weapons, blinding lasers, land mines, cluster munitions, using the environment as a weapon, placing weapons in space, or certain delivery mechanisms or deployment postures of nuclear weapons. Many other attempts at arms control have failed, from the papal decrees denouncing the use of the crossbow in the Middle Ages to 20th-century attempts to ban aerial attacks on cities, regulate submarine warfare, or eliminate nuclear weapons. The United Nations began a series of meetings in 2014 to discuss the perils of autonomous weapons. But so far the progress has been far slower than the pace of technological advances. Despite that lack of success, a growing number of voices have begun calling for a ban on autonomous weapons. Since 2013, 76 nongovernmental organizations across 32 countries have joined a global Campaign to Stop Killer Robots. To date, nearly 4,000 artificial intelligence and robotics researchers have signed an open letter calling for a ban. More than 25 national governments have said they endorse a ban, although none of them are major military powers or robotics developers. But such measures only tend to succeed when the weapons in question are of marginal value, are widely seen as especially horrific or destabilizing, are possessed by only a few actors, are clearly distinguished from other weapons, and can be easily inspected to verify disarmament. None of these conditions applies to autonomous weapons. Even if all countries agreed on the need to restrain this class of arms, the fear of what others might be doing and the inability to verify disarmament could still spark an arms race. Less ambitious regulations could fare better, such as a narrow ban on anti-personnel autonomous weapons, a set of rules for interactions between autonomous weapons, or a broad principle of human involvement in lethal force. While such modest efforts might mitigate some risks, however, they would leave countries free to develop many types of autonomous weapons that could still lead to widespread harm. Humanity stands at the threshold of a new era in war, in which machines will make life-or-death decisions at speeds too fast for human comprehension. The risks of such a world are real and profound. Autonomous weapons could lead to accidental death and destruction at catastrophic scales in an instant. The unrestrained pursuit of fully autonomous weapons could lead to a future where humans cede control over what happens on the battlefield, but the critical decisions about how this technology is used still rest in human hands.

**An Autonomous Incidents Agreement would build trust and avoid escalating crises or conflicts – empirically prove at sea.**

**Horowitz and Kahn, 2021 – Senior and Research Fellows at the Council on Foreign Relations** [Michael and Lauren, The Washington Quarterly 3-19-2021 “Leading in Artificial Intelligence through Confidence Building Measures”  [https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/0163660X.2021.2018794 acc on 6-21-2022](%20https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/0163660X.2021.2018794%20acc%20on%206-21-2022%20) TM]

As described above, one specific concern about AI-enabled autonomous systems is the way their behavior on the battlefield could inadvertently be escalatory due to accidents and unanticipated interactions between adversarial autonomous systems. A key element driving these potential escalation dynamics is uncertainty about how AI-enabled autonomous systems will perform—including whether that behavior will be explainable—and a lack of familiarity by militaries regarding how to react. An Autonomous Incidents Agreement, borrowing from the US-Soviet Incidents at Sea Agreement during the Cold War, could reduce these risks in peacetime and early on in crises.32 The 1972 US-Soviet Incidents at Sea Agreement, created as part of the SALT I process, addressed ongoing US-Soviet naval interactions that some feared could lead to inadvertent escalation. These interactions included high-speed surveillance, accidental firing, unanticipated naval movements, and simulated attacks. Notification procedures and information sharing on naval activities decreased the risk of accidents and inadvertent escalation.33 The agreement would not have been sufficient to prevent war if either side wanted it, but in peacetime or early in a crisis, it could help differentiate normal from unusual behavior in ways that could increase reassurance and establish guides for what activity was within expectations. An Autonomous Incidents Agreement (AIA) could potentially serve a similar purpose between the United States and China as well as other countries. An AIA could become especially relevant early in the AI age, when there is more uncertainty about how AI-enabled capabilities will function and how or if they will disrupt existing dynamics and tacitly accepted military movements and actions. Elements of an AIA could include notification provisions when deploying AI enabled autonomous systems. Information sharing about broad system parameters, such as programming uninhabited sea vehicles to stay a certain distance away from other ships, could also help make it clearer when a potentially escalatory incident is, in fact, an accident. By focusing explicitly on avoiding accidents and inadvertent escalation, an Autonomous Incidents Agreement could therefore serve as a mechanism for cooperation, even among competitors like the United States and China. At best, it could represent a building block for cooperation in preventing unintentional conflict. At worst, if the United States proposed such a CBM and China turned it down, it would illustrate to the international community America’s interest in military AI cooperation, thereby increasing US credibility on these issues with allies and partners.

### --Extend – Military Technology Net Benefit

#### The counterplan solves for AI safety without revealing TEVV secrets to Russia and China.

**Horowitz and Kahn, 2021 – Senior and Research Fellows at the Council on Foreign Relations** [Michael and Lauren, The Washington Quarterly 3-19-2021 “Leading in Artificial Intelligence through Confidence Building Measures”  [https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/0163660X.2021.2018794 acc on 6-21-2022](%20https://www-tandfonline-com.proxy.lib.umich.edu/doi/full/10.1080/0163660X.2021.2018794%20acc%20on%206-21-2022%20) TM]

As noted above, one might be opposed to dialogue on AI safety with China and Russia, fearing that it could reveal information that China and Russia would use to make their military applications of AI more reliable, meaning they would be more effective. However, these proposed discussions would not include methods to make algorithms more reliable or divulge details on how TEVV works. Dialogues of that nature might provide China and Russia with a roadmap to make their algorithms more effective. Instead, dialogue would focus on ideas about strategic stability and areas of mutual interest such as reducing the chance of inadvertent escalation. Communicating the importance the United States places on safety and reliability and emphasizing already public information on US safety checks before deploying weapons could also generate the advantages described above.

### Nuclear Human Control Counterplan – 1NC

#### Text – The United States should increase security cooperation with NATO to require positive human control over all nuclear launch decisions

**Counterplan solves the risk of nuclear accidents by requiring human control**

**Horowitz and Scharre 2021 - Director of the Emerging Capabilities Policy Office in the Office of the Under Secretary of Defense for Policy and  Vice President and Director of Studies at CNAS** [Michael, Paul, January 12 2021, “AI and International Stability: Risks and Confidence- Building Measures”, <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>, Acc 6/18/22. M.A.]

Specific Mission-Related CBMs: Nuclear Operations The integration of AI, autonomy, and/or automation into nuclear command-and-control, early warning, and delivery systems poses unique risks to international stability because of the extreme consequences of nuclear accidents or misuse.[71](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn71) One option for mitigating these risks could be for nations to set limits on the integration of AI, autonomy, or automation into their nuclear operations. Some U.S. military leaders and official DoD documents have expressed skepticism about integrating uninhabited vehicles into plans surrounding nuclear weapons. The Air Force’s 2013 Remotely Piloted Aircraft (RPA) Vector report proposed that nuclear strike “may not be technically feasible unless safeguards are developed and even then may not be considered for [unmanned aircraft systems] operations.”[72](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn72) U.S. Air Force general officers have been publicly skeptical about having uninhabited vehicles armed with nuclear weapons. General Robin Rand stated in 2016, during his time as head of Air Force Global Strike Command, that: “We’re planning on [the B-21] being manned. … I like the man in the loop … very much, particularly as we do the dual-capable mission with nuclear weapons.”[73](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn73) Other U.S. military leaders have publicly expressed support for limits on the integration of AI into nuclear command-and-control. In September 2019, Lieutenant General Jack Shanahan, head of the DoD Joint AI Center, said, “You will find no stronger proponent of the integration of AI capabilities writ large into the Department of Defense, but there is one area where I pause, and it has to do with nuclear command and control.” In reaction to the concept of the United States adopting a “dead hand” system to automate nuclear retaliation if national leadership were wiped out, Shanahan said, “My immediate answer is ‘No. We do not.’ … This is the ultimate human decision that needs to be made which is in the area of nuclear command and control.”[74](https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures#fn74) While the motivation for these statements about limits on the use of autonomy may or may not be strategic stability—bureaucratic factors could also be at play—they are examples of the kinds of limits that nuclear powers could agree to set, unilaterally or collectively, on the integration of AI, autonomy, and automation into their nuclear operations. Nuclear states have a range of options for how to engage with these kinds of risks. On one end of the spectrum are arms control treaties with some degree of verification or transparency measures to ensure mutual trust in adherence to the agreements. On the other end of the spectrum are unilateral transparency measures, which could have varying degrees of concreteness ranging from informal statements from military or civilian leaders along the lines of Shanahan’s and Rand’s statements, all the way to formal declaratory policies. In between are options such as mutual transparency measures, statements of principles, or non-legally binding codes of conduct or other agreements between nuclear states to ensure human control over nuclear weapons and nuclear launch decisions. Even if states that desired these restraints found themselves in a position where others were unwilling to adopt more binding commitments, there may be value in unilateral transparency measures both to reduce the fears of other states and to promulgate norms of responsible state behavior. As with other areas, it is important to consider incentives for defection from an agreement and the extent to which one state’s voluntary limitations depend on verifying others’ compliance with an agreement. If some states, such as the United States, desire strict positive human control over their nuclear weapons and nuclear launch authority for their own reasons, then verifying others’ behavior, while desirable, may not be a necessary precondition to those states adopting their own limits on the use of AI, autonomy, or automation in nuclear operations. Two possible CBMs for AI applications in the nuclear arena involve nuclear weapons states agreeing to strict human control over nuclear launch decisions and ensuring any recoverable delivery vehicles are human-inhabited, to ensure positive human control. Strict Human Control over Nuclear Launch Decisions One CBM for uses of AI in the nuclear arena would involve an agreement by nuclear powers to ensure positive human control over all nuclear launch decisions. This type of agreement would preclude automated “dead hand” systems or any other automatic trigger for the use of nuclear weapons. The benefit of such a CBM would be to reduce the risk of accidental nuclear war. It would preclude a machine malfunction leading directly to the use of nuclear weapons without a human involved in the process. Agreement on positive human control over nuclear launch decisions could also be a mechanism for dialogue with newer nuclear powers, helping generate more transparency over their nuclear launch decisions.

### Civilian Counterplan Links

**International cooperation on Civilian AI research strengthens relationships – serves a foundation for future agreements**

**Horowitz and Scharre 2021 - Director of the Emerging Capabilities Policy Office in the Office of the Under Secretary of Defense for Policy and  Vice President and Director of Studies at CNAS** [Michael, Paul, January 12 2021, “AI and International Stability: Risks and Confidence- Building Measures”, <https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures>, Acc 6/18/22. M.A.]

International efforts to promote shared civilian research on AI safety could be a low-level CBM that would not explicitly involve military action. Shared civilian research would build scientific cooperation between nations, which could serve as a building block for overall cooperation. Focusing cooperation on AI safety, an area of shared interest, might also make more nations willing to sign on to participate. An analogy to this in the U.S.-Soviet context is the Apollo-Soyuz mission in 1975, whose intent was to promote cooperation between civilian scientists on a shared agenda. Similarly, nations could work to foster increased cooperation and collaboration between civilian scientists on AI safety.

### Prize Counterplan Links

**Performance prizes incentivize private companies to solve AI engineering problems – government regulations cannot solve due to uncertain funding and goals.**

**Aronhime and Cocron, 2021 -Professors of Engineering at Johns Hopkins University** [Lawrence and Alexander NATO Review July 19 ‘NATO’s Innovation Challenge’ https://www.nato.int/docu/review/articles/2021/07/19/natos-innovation-challenge/index.html, BK]

We may now find ourselves in similar times. The Alliance faces rising competitors and, like the space race before it, the competition will require innovation in emerging and highly disruptive technologies. At the London Leaders’ meeting in 2019, the Alliance defined seven areas of emerging and potentially disruptive technologies that will be “highly influential for the development of future military capabilities”. They are data, Artificial Intelligence (AI), autonomy, space, hypersonics, quantum and biotechnology. For policymakers, the question immediately arises: “Is the Alliance ahead or behind?” To answer such a question properly, the terms need to be framed in such a way that an answer can be measured and thus either falsified or proved. For example, has a member of the Alliance developed a vehicle that can safely travel at Mach 10 from New York to Paris? An answer to this question is measurable and falsifiable. The question does not ask how such a vehicle should be produced or what the design of its architecture and components should look like. Or how many patents and engineers are needed. Rather, the question has a clearly stated goal that can be used to challenge and inspire government, industry and academia, much as President Kennedy did. Sputnik moments In an era of constrained budgets, the Alliance must make its innovation investments wisely in order to avoid future “Sputnik moments.” But such moments become increasingly likely if it is not clear where the Alliance is behind or where it is ahead of its competitors. The approach thus far has been to track national inputs: how many patents filed, how many science and technology researchers employed, how much research and development money invested, how many new technology startups founded and funded. In other words, invest and hope for the best. These input metrics certainly have their use, but they do not reliably predict which countries will innovate in potentially important ways. Would looking at these input measures have enabled the Alliance to predict the Soviet Union’s successful Sputnik launch? In fact, innovation in the space race was propelled by clear performance goals: to put a satellite safely into orbit, then a live animal, then a human, and so forth. In one of the most innovative periods in world history, there was no question who was ahead and who was behind. This experience can help us think more clearly about today. It was by setting specific performance challenges that innovation was driven relentlessly forward. The US president set the ultimate objective: to go to the moon and back. This created a cascade of intermediate milestones and sub-challenges: get astronauts into orbit, have them manoeuvre an orbiting spacecraft, do orbital rendezvous, perform extra vehicular activity, etc. Each of these were necessary conditions for getting to the moon, and, just as importantly, getting back. Great innovations start with great problems Instead of investing and hoping for the best, the Alliance could focus on setting audacious performance goals, and then letting the triple helix get to work. For the most pressing or significant problems, the Alliance could create formal challenges with prizes and incentives to encourage innovators to give it their best shot. This approach has worked well in the past and is currently used by a few industrial companies, government agencies, and private philanthropists to stimulate innovation. The Ansari X prize encouraged the aircraft designer Burt Rutan to successfully develop a reusable commercial space plane. In the 1920’s, the Orteig Prize encouraged a little-known airmail pilot to develop a specialized airplane and fly it solo across the perilous North Atlantic. Charles Lindbergh probably would not have risked his life if there were no prize or glory waiting for him at the other end. This is perhaps the most famous historical example (and one that, in the spirit of the Alliance, features an allied aviator bravely solving a transatlantic problem). There are many more.

### Accepted Risk Counterplan Links

**Allowing an increase in the acceptance of self risk for autonomous weapons cuts down on accidents – it allows more time for reaction**

**Atherton 2022 – Military Technology Journalist** [Kelsey, 5/6/22, “Understanding the errors introduced by military AI applications”, <https://www.brookings.edu/techstream/understanding-the-errors-introduced-by-military-ai-applications/>, 6/18/22, LND]

Aware of the potential for error, one way to adopt autonomous systems while addressing the risk to civilians and servicemembers is to shift toward a posture in which risk is borne primarily by the machine. The 2003 shootdowns involved Patriot missiles acting in self-defense and misidentifying their enemy. By accepting greater risk to autonomous systems—that they might be destroyed or disabled—autonomous systems can avoid the risk of friendly fire or civilian casualties by “using tactical patience, or allowing the platform to move in closer to get a more accurate determination of whether a threat actually exists,” as Larry Lewis, the author of the 2019 CNA report, argues. Rather than quickly firing in self-defense, this view argues for patience and sacrificing a measure of speed in favor of accuracy.

More broadly, Lewis recommends a risk management approach to using AI. While the specific nature of every given error is hard to anticipate, the range of bad and undesired outcomes can fall in similar categories of error or outcome. Planning for AI incorporated into weapons, sensors, and information displays could include an awareness of error, and present that information in a useful way without adding to the cognitive load of the person using the machine.

Artificial Intelligence has already moved beyond the speculative to tangible, real-world applications. It already informs the targeting decisions of military weapons, and will increasingly shape how people in combat use machines and tools. Adapting to this future, as the Pentagon and other military establishments seem intent to do, means planning for error, accidents, and novel harm, the way militaries have already adapted to such error in human hands.