Texas National Security Review: Volume 4, Issue 3 (Summer 2021)
Print: ISSN 2576-1021 Online: ISSN 2576-1153

DEBUNKING THE AI ARMS RACE THEORY

Paul Scharre

There is no AI arms race. However, military competition in AI does still pose certain risks. These include losing human control and the acceleration of warfare, as well as the risk that perceptions of an arms race will cause competitors to cut corners on testing, leading to the deployment of unsafe AI systems.

In 2015, a group of prominent AI and robotics researchers signed an open letter warning of the dangers of autonomous weapons. "The key question for humanity today," they wrote, "is whether to start a global AI arms race or to prevent it from starting. If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable."[1] Today, many nations are working to apply AI for military advantage, and the term "AI arms race" has become a catchphrase used by both critics and proponents of AI militarization. In 2018, then-Under Secretary of Defense Michael Griffin, calling for the United States to invest more in AI, stated, "There might be an artificial intelligence arms race, but we're not yet in it."[2] In a 2020 *Wired* article, Will Roper, then chief acquisition officer for the U.S. Air Force, warned of the risks of falling behind in a "digital arms race with China."[3]

The so-called AI arms race has become a common feature in news headlines,[4] but the arms race framing fails to match reality. While nations are clearly competing to develop and adopt AI technology for military use, the character of that competition does not meet the traditional definition of an arms race. Military AI competition nevertheless does pose risks. The widespread adoption of military AI could cause warfare to evolve in a manner that leads to less human control and to warfare becoming faster, more violent, and more challenging in terms of being able to manage escalation and bring a war to an end. Additionally, perceptions of a "race" to field AI systems before competitors do could cause nations to cut corners on testing, leading to the deployment of unsafe AI systems that are at risk of accidents that could cause unintended escalation or destruction. Even if fears of an "AI arms race" are overblown, military AI competition brings real risks to which nations should attend. There are concrete steps nations can take to mitigate some of these dangers.

Current Military AI Competition Is Not an "Arms Race"

As Heather Roff has written, the arms race framing "misrepresents the competition going on among countries."[5] To begin with, AI is not a weapon. AI is a general-purpose enabling technology with myriad applications. It is not like a missile or a tank. It is more like electricity, the internal combustion engine, or computer networks.[6] General-purpose technologies like AI have applications across a range of industries. *Wired* magazine co-founder Kevin Kelly has argued that it "will enliven inert objects, much as electricity did more than a century ago. Everything that we formerly electrified we will now cognitize."[7]

Nations may very well be in a *technology race* to adopt AI across a range of industries. AI will help to improve economic productivity and, by extension, economic and military power. During the industrial revolution, early adopters of industrial technology

1 "Autonomous Weapons: An Open Letter from AI & Robotics Researchers," Future of Life Institute, 2015, https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1.

2 Brandon Knapp, "DoD Official: US Not Part of AI Arms Race," *C4ISRNET*, April 10, 2018, https://www.c4isrnet.com/it-networks/2018/04/10/dod-official-us-not-part-of-ai-arms-race/.

3 Will Roper, "There's No Turning Back on AI in the Military," *Wired,* Oct. 24, 2020, https://www.wired.com/story/opinion-theres-no-turning-back-on-ai-in-the-military/.

4 Andrew Imbrie, et al., "Mainframes: A Provisional Analysis of Rhetorical Frames in AI," Center for Security and Emerging Technology, August 2020, https://cset.georgetown.edu/research/mainframes-a-provisional-analysis-of-rhetorical-frames-in-ai/.

5 Heather M. Roff, "The Frame Problem: The AI "arms race" isn't one," Bulletin of the Atomic Scientists, April 29, 2019, https://thebulletin.org/2019/04/the-frame-problem-the-ai-arms-race-isnt-one/.

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

7 Kevin Kelly, "The Three Breakthroughs that Have Finally Unleashed AI on the World," *Wired,* Oct. 27, 2014, https://www.wired.com/2014/10/future-of-artificial-intelligence/.

significantly increased their national power. From 1830 to 1890, Britain and Germany, which were both early industrializers, more than doubled their per capita gross national product while Russia, which lagged in industrialization, increased its per capita gross national product by a mere 7 percent over that 60-year period.[8] These technological advantages led to increased economic and military power, most notably for Europe relative to the rest of the world. In 1790, Europe (collectively), China, and India (including what is now Pakistan and Bangladesh) held roughly the same shares of global manufacturing output, with Europe and India each holding about one-quarter of global manufacturing output and China holding roughly one-third. They all had approximately equivalent levels of per capita industrialization at that time. But the industrial revolution skyrocketed European economic productivity. By 1900, Europe collectively controlled 62 percent of global manufacturing output, while China held only six percent and India less than two percent. These economic advantages translated into military power. By 1914, Europeans occupied or controlled over 80 percent of the world's land surface.[9]

Being ahead of the curve in adopting AI is likely to lead to significant national advantages. Although AI can increase military capabilities, the more consequential advantages over the long term may come from non-military AI applications across society. Long-term benefits from AI could include increased productivity, improved healthcare outcomes, economic growth, and other indicators of national well-being. Increasing productivity is especially significant because it has a compounding effect on economic growth. Over the long term, technological progress is the main driver of economic growth.[10]

Of course, AI can also be used for weapons. Militaries around the world are actively working to adopt AI to improve their military capabilities. Yet the militarization of AI does not, at present, meet the traditional definition of an arms race, despite the rhetorical urgency of many national leaders. Michael D. Wallace, in his 1979 article "Arms Races and Escalation," defined an arms race as "involving simultaneous abnormal rates of growth in the military outlays of two or more nations" resulting from "the competitive pressure of the military itself, and not from domestic forces exogenous to this rivalry." Wallace further stated that the concept of an arms race only applied "between nations whose foreign and defense policies are heavily interdependent" and who have "roughly comparable" capabilities.[11] AI is being adopted by many countries around the globe.[12] Arguably at least some of the dyads, such as the United States and China, meet Wallace's definition in terms of being nations with "roughly comparable" capabilities, locked in competition, "whose foreign and defense policies are heavily interdependent." However, AI fails the arms race test in the critical area of spending.

Wallace distinguished arms races from the normal behavior of states to improve their military forces. A state that adopts a new technology and modernizes its military forces is not automatically in an arms race, under Wallace's definition, even if the modernization is aimed at competition with another country. The decisive factor in qualifying as an arms race, according to Wallace, is the rate of growth in defense spending. Wallace characterized arms races as resulting in abnormally large growth rates in defense spending, beyond the historical average of 4 to 5 percent annual growth (in real dollars). In an arms race, annual growth rates are above 10 percent or even as high as 20 to 25 percent.[13] Other scholars define arms races using different quantitative thresholds — and some definitions lack clear quantitative thresholds at all — but the existence of rapid increases in defense spending or military forces above normal levels is a common criterion in the scholarly literature on arms races.[14]

Arms races result in situations in which two or more countries are locked in spiraling defense spending, grabbing ever-greater shares of national

8 Paul Kennedy, *The Rise and Fall of the Great Powers: Economic Change and Military Conflict from 1500 to 2000* (New York: Random House, 1987), 171.

9 Kennedy, *The Rise and Fall of the Great Powers,* 149–50.

10 YiLi Chien, "What Drives Long-Run Economic Growth?" St. Louis Fed On the Economy Blog, June 1, 2015, https://www.stlouisfed.org/on-the-economy/2015/june/what-drives-long-run-economic-growth.

11 Michael D. Wallace, "Arms Races and Escalation," *Journal of Conflict Resolution* 23 no. 1 (March 1979), 5, https://www.jstor.org/stable/173649.

12 Tim Dutton, "An Overview of National AI Strategies," *Medium*, June 28, 2018, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

13 Wallace, "Arms Races and Escalation," 6.

14 For example, see Colin Gray, "The Arms Race Phenomenon," *World Politics* 24, no. 1 (October 1971), 41, https://doi.org/10.2307/2009706; Theresa Clair Smith, "Arms Race Instability and War," *Journal of Conflict Resolution* 24, no. 2 (June 1980): 255–56, https://doi.org/10.1177%2F002200278002400204; Michael F. Altfeld, "Arms Races?—And Escalation? A Comment on Wallace," *International Studies Quarterly* 27, no. 2 (June 1983): 225–26, https://doi.org/10.2307/2600547; Paul F. Diehl, "Arms Races and Escalation: A Closer Look," *Journal of Peace Research* 20, no. 3 (September 1983), 206–08, https://www.jstor.org/stable/423792; and Toby J. Rider, Michael G. Findley, and Paul F. Diehl, "Just Part of the Game? Arms Races, Rivalry, and War," *Journal of Peace Research* 48, no. 1 (January 2011): 90, https://www.jstor.org/stable/29777471.

treasure often with little to no net gain in relative advantage over the other. Classic historical examples include the Anglo-German naval arms race prior to World War I and the U.S.-Soviet nuclear arms race during the Cold War. Military AI spending today clearly does not meet these criteria of abnormally large growth rates in defense spending. AI defense spending is difficult to calculate due to the general-purpose nature of AI technology. Unlike ships or ballistic missiles, AI systems cannot be easily counted. Nevertheless, even crude estimates of defense spending show that military AI investments are nowhere near large enough to constitute an arms race. An independent estimate by Bloomberg Government of U.S. defense spending on AI identified $5 billion in AI-related research and development in fiscal year 2020, or roughly 0.7 percent of the Department of Defense's over $700 billion budget.[15] The scale of military AI spending, at least at present, is nowhere near large enough to warrant the title of "arms race." (Adding in private sector spending, which constitutes the bulk of AI investment, would lead to larger figures but would further belie the claim of an "arms" race since most private sector AI investment is not in weapons.)

AI Competition and the Security Dilemma

Even if military AI spending does not rise to the level of an "arms race," many nations are nevertheless engaged in a security competition in the adoption of military AI, a competition that does pose risks. The situation that states find themselves in with regard to AI competition is much more accurately described as a security dilemma,[16] a more generalized competitive dynamic between states than the more narrowly defined "arms race." In his 1978 article, "Cooperation Under the Security Dilemma," Robert Jervis defined the security dilemma as follows: "[M]any of the means by which a state tries to increase its security decrease the security of others."[17] As Charles Glaser has pointed out, it is not obvious from this definition why it would be intrinsically bad for an increase in one state's security to come at the expense of another's security.[18] In fact, decreasing the security of other states could have beneficial effects in enhancing deterrence and reducing the risks of aggression or achieving a favorable balance of power in a region, which could lead to greater political influence. The problem comes in the second- and third-order effects that could develop when another state reacts to having its security reduced. Responses could include counterbalancing with a net effect of no change in security (or worsening security). Glaser argues that there are some situations in which security competition is a rational strategy for a state to pursue even if competitors will arm in response. In other situations, arming may be a suboptimal strategy for a state, which would be better served by restraint or pursuing arms control.[19]

Security competition could even leave both states worse off than before. This can occur during a traditional arms race if nations expend vast sums of money in an unsuccessful attempt to gain an advantage over one another, with the result that both nations divert funds from non-defense expenditures. If the outcome of a security competition is the same relative military balance as before, the balance of power may not have meaningfully changed, but both nations could face diminished economic and social well-being at home relative to if they had avoided a security competition. Even absent this "guns vs. butter" tradeoff, however, there are other ways in which security competition can lead to a net negative outcome for both states.

One way this could occur is if military innovation and the development of new capabilities alter the character of warfare in a manner that is more harmful, more destructive, less stable, or otherwise less desirable than before. In his 1997 article, "The Security Dilemma Revisited," Glaser gave the example of military capabilities that shifted warfare to a more offense-dominant regime.[20] There are other ways in which warfare could evolve in a net negative direction as well. For example, in World War I, Germany's interest in developing and deploying chemical weapons was spurred in part due

15 "Artificial Intelligence Index Report 2021," Stanford University Human-Centered Artificial Intelligence, https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf, 168.

16 John H. Herz, "Idealist Internationalism and the Security Dilemma," *World Politics* 2, no. 2 (January 1950): 157–80, https://doi.org/10.2307/2009187; and John H. Herz, *Political Realism and Political Idealism* (Chicago: University of Chicago Press, 1951).

17 Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30, no. 2 (January 1978): 169, https://doi.org/10.2307/2009958.

18 Charles L. Glaser, "The Security Dilemma Revisited," *World Politics* 50, no. 1 (October 1997): 174, https://www.jstor.org/stable/25054031.

19 Charles L. Glaser, "When Are Arms Races Dangerous?" *International Security* 28, no. 4 (Spring 2004): 44–84, https://doi.org/10.1162/0162288041588313.

20 Glaser, "The Security Dilemma Revisited," 176.

to fears about France's developments in poison gas.[21] The result was the introduction of a weapon that increased combatant suffering on both sides, without delivering a significant military advantage to either. The same could occur with AI: It could alter the character of warfare in a way that would be a net negative for all participants.

One state's pursuit of greater automation and faster reaction times undermines other states' security and leads them to similarly pursue more automation just to keep up.

An Accelerating Tempo of Warfare

One possibility for how AI could alter warfare in a manner that would leave all states worse off would be if it accelerated the tempo of war past the point of human control, making warfare faster, more violent, and less controllable. There are advantages to adding intelligence into machines, but given the limitations of AI systems today, the optimal model for achieving the highest quality decision-making would be a joint human-machine architecture that combines human and machine decision-making. One way in which machines outperform humans, however, is in speed. It is possible to envision a competitive dynamic in which countries feel compelled to automate increasing amounts of their military operations in order to keep pace with adversaries. Then-Deputy Secretary of Defense Robert O. Work summed up the dilemma when he asked, "If our competitors go to Terminators and we are still operating where the machines are helping the humans and it turns out the Terminators are able to make decisions faster, even if they're bad, how would we respond?"[22] This is a classic security dilemma. One state's pursuit of greater automation and faster reaction times undermines other states' security and leads them to similarly pursue more automation just to keep up.

If states fall victim to this trap, it could lead to all states being less secure, since the pursuit of greater automation would not merely be an evolution in weapons and countermeasures that simply leads to the creation of new weapons in the future. At some point, warfare could shift to a qualitatively different regime in which humans have less control over lethal force as decisions become more automated and the accelerating tempo of operations pushes humans "out of the loop" of decision-making. Some Chinese scholars have hypothesized about a battlefield "singularity," in which the pace of combat eclipses human decision-making.[23] U.S. scholars have used the term "hyperwar" to refer to a similar scenario.[24] While the speed of engagement necessitates automation in some limited areas today, such as immediate localized defense of ships, bases, and vehicles from rocket and missile attack, expanding this zone of machine control into broader areas of war would be a significant development. Less human control over warfare could lead to wars that are less controllable and that escalate more quickly or more widely than humans intend. Similarly, limiting escalation or terminating conflicts could be more challenging if the pace of operations on the battlefield exceeds human decision-making. Political leaders would have a command-and-control problem in which their military forces are operating "inside" (i.e., faster than) their own decision cycle. The net effect of the quite rational desire for nations to gain an edge in speed could lead to an outcome that is worse for all. Yet, competitive dynamics could nevertheless drive such a result.

Financial markets provide an example of this dynamic in a non-military competitive environment. Automation introduced into financial markets, especially high-frequency trading in which trades are executed at super-human speeds in milliseconds, has contributed to unstable market conditions that can lead to "flash crashes," in which prices rapidly and dramatically shift.[25] Financial regulators have

21 Charles E. Heller, "Chemical Warfare in World War I: The American Experience, 1917–1918," *Leavenworth Papers,* No. 10, Combat Studies Institute, U.S. Army Command and General Staff College, September 1984, 6, https://apps.dtic.mil/dtic/tr/fulltext/u2/a189331.pdf.

22 Robert O. Work, "Remarks at the Atlantic Council Global Strategy Forum," Washington, DC, May 2, 2016, http://www.atlanticcouncil.org/events/webcasts/2016-global-strategy-forum.

23 Elsa B. Kania, "Battlefield Singularity," Center for a New American Security, Nov. 28, 2017, https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power; and Chen Hanghui [陈航辉], "Artificial Intelligence: Disruptively Changing the Rules of the Game [人工智能：颠覆性改变"游戏规则]," China Military Online, March 18, 2016, http://www.81.cn/jskj/2016-03/18/content_6966873_2.htm. Chen Hanghui is affiliated with the Nanjing Army Command College.

24 John R. Allen and Amir Husain, "On Hyperwar," *Proceedings,* July 2017, https://www.usni.org/magazines/proceedings/2017/july/hyperwar.

25 *Findings Regarding the Market Events of May 6, 2010,* U.S. Commodity Futures Trading Commission and U.S. Securities and Exchange Commission, Sept. 30, 2010, 2, http://www.sec.gov/news/studies/2010/marketevents-report.pdf; and Maureen Farrell, "Mini Flash Crashes: A Dozen a Day," *CNN,* March 20, 2013, http://money.cnn.com/2013/03/20/investing/mini-flash-crash/index.html.

responded by employing "circuit breakers" that automatically halt trading for a pre-determined period of time if the price moves too quickly.[26] Financial markets have the benefit of a regulator who can force cooperative measures on competitors to address suboptimal outcomes. Under conditions of anarchy in the international security environment, any such cooperation would have to come from states themselves.

The dynamic of a competition in speed is *like* an arms race, if we expand the definition of an arms race to be more in line with biological examples of competitive co-evolution. Biologists often use the metaphor of an arms race to explain "an unstable runaway escalation" of adaptation and counter-adaption that can occur in animals.[27] This can occur between species, such as predator and prey, or within species, such as males evolving in competition for females. Biological arms races can manifest in a variety of ways, such as competitions between predator and prey with regard to camouflage vs. detection and armor vs. claws, as well as speed, cognitive abilities, poison, deception, or other attributes that might increase chances of survival.[28] This broader biological definition of an arms race is more in line with the potential for an escalating "arms race in speed" among nations that leads to greater automation in warfare. While this concept does not meet the traditional definition of an arms race in the security studies literature, it is nevertheless a useful concept to describe the potential for a co-evolution in speed that leads to no net relative advantage and in fact may leave both sides worse off.

Race to the Bottom on Safety

A related risk of a "racing" dynamic among competitors could come from an acceleration, not of the pace of operations on the battlefield, but of the process of fielding new AI systems. AI systems today have a host of safety and security problems that can make them brittle, unreliable, and insecure.[29] Because machine learning in particular can create new ways in which systems can fail, militaries face novel challenges in adopting AI systems.[30] Militaries will have to adopt new methods to test, evaluate, verify, and validate AI systems (also known as TEVV).[31] Such concerns related to autonomy are well known in the U.S. defense community,[32] although at present they have not been solved to a satisfactory degree. Machine learning introduces additional challenges with regard to testing, evaluation, verification, and validation. A rush to field AI systems before they are fully tested could result in a "race to the bottom" on safety, with militaries fielding accident-prone AI systems.

There are strong bureaucratic and institutional imperatives for militaries to field systems that are robust and secure. Indeed, designing systems to military specification standards often means making them more robust for a wider range of environmental conditions and shocks than comparable commercial systems, even at the expense of other aspects of performance, such as size, weight, or usability. AI presents novel challenges, however, in achieving the robustness needed for operating in the complex, hazardous, and adversarial environments that often characterize military operations.

Certain AI methods today, such as deep learning, remain relatively immature with significant reliability challenges. A 2017 Department of Defense report by the JASON scientific advisory group explained that deep neural networks

> are immature as regards the "illities", including reliability, maintainability, accountability, validation and verification, debug-ability, evolvability, fragility, attackability, and so forth. … Further, it is not clear that the existing AI paradigm is immediately amenable to any sort of software engineering validation and verification. This is a serious

26 *Investor Bulletin: Measures to Address Market Volatility*, U.S. Securities and Exchange Commission, July 1, 2012, https://www.sec.gov/oiea/investor-alerts-bulletins/investor-alerts-circuitbreakersbulletinhtm.html. Specific price bands are listed here: Jason Fernando, "Circuit Breaker," Investopedia, Nov. 18, 2003, updated Feb. 26, 2021, http://www.investopedia.com/terms/c/circuitbreaker.asp.

27 Richard Dawkins and John Richards Krebs, "Arms Races Between and Within Species," *Proceedings of the Royal Society of London. Series B, Biological Sciences* 205, no. 1161 (September 21, 1979): 489, https://doi.org/10.1098/rspb.1979.0081.

28 Dawkins and Krebs, "Arms Races Between and Within Species," 489–98.

29 Dario Amodei, et al., "Concrete Problems in AI Safety," arXiv 1606.06565 (2016), https://arxiv.org/abs/1606.06565; and Ram Shankar, et al., "Failure Modes in Machine Learning," Microsoft, Nov. 11, 2019, https://docs.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning.

30 *Perspectives on Research in Artificial Intelligence and Artificial General Intelligence Relevant to DoD*, The MITRE Corporation January 2017, https://fas.org/irp/agency/dod/jason/ai-dod.pdf; and Danielle C. Tarraf, et al., *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations* (Santa Monica, CA: RAND Corp., 2019),https://www.rand.org/pubs/research_reports/RR4229.html.

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

32 "Autonomous Horizons: System Autonomy in the Air Force – A Path to the Future, Volume 1: Human-Autonomy Teaming," United States Air Force Office of the Chief Scientist, June 2015, 23, https://www.af.mil/Portals/1/documents/SECAF/AutonomousHorizons.pdf.

> issue, and is a potential roadblock to DoD's [Department of Defense's] use of these modern AI systems, especially when considering the liability and accountability of using AI in lethal systems.[33]

The Defense Department's 2018 AI strategy calls for building AI systems that are "resilient, robust, reliable, and secure."[34] Yet, the current state of technology makes achieving this goal particularly difficult for AI systems that incorporate deep learning, a subfield of AI that has seen significant growth and attention in recent years. While there is active research underway to improve AI safety and security, militaries will have to adapt to the technology as it currently is, at least for the time being. An ideal process would be for militaries to engage in experimentation, prototyping, and concept development, but also to subject AI systems to rigorous TEVV under realistic operational conditions before deployment. Taking shortcuts on testing and evaluation and fielding a system before it is fully tested could lead to accidents, which, in some settings, could undermine international stability.

In evaluating new technologies, militaries may be relatively accepting of the risk of accidents, which may lead them to tolerate the deployment of systems that have reliability concerns. In building and fielding new capabilities, militaries have to weigh the possibility of an accident occurring against other concerns, such as forgoing valuable military capabilities. The military operational environment is fraught with risk, in both training and real-world operations. Military institutions balance managing this risk with other factors, such as the need for training, developing new capabilities, or accomplishing the mission. Military institutions view casualties from training accidents or testing new capabilities as a tragic but unavoidable part of the business of preparing for war. Militaries expect high performance from their forces, often while they are performing dangerous tasks, but militaries neither demand nor expect accident-free operations in most settings.[35] From 2006 to 2020, over 5,000 U.S. servicemembers were killed in non-war related accidents, the majority of which occurred within the United States. Accidents overall accounted for nearly 32 percent of U.S. servicemember deaths during this period, and even accounted for a significant portion of servicemember deaths in Iraq (19 percent) and Afghanistan (16 percent).[36] These accident rates are not unusual for the U.S. armed forces. This is business as usual. Accidents draw the attention of senior military and civilian officials when a spate of accidents occur in a short amount of time — such as a series of aircraft crashes,[37] ship collisions,[38] or training accidents.[39] Yet, as one report on naval accidents from 1945 to 1988 notes, "peacetime naval accidents are a fact of life."[40] The same is true of military air and ground operations. Other nations' militaries may do an even poorer job of managing risk when it comes to accidents than the U.S. military. For example, the Soviet/Russian submarine community has a much higher accident rate than the U.S. submarine community.[41]

New technologies in particular present an increased risk of accidents, yet militaries may press ahead out of a desire to develop and field what they perceive to be a valuable capability. For example, the V-22 Osprey tiltrotor aircraft suffered four crashes during development, killing 30 U.S. servicemembers in total, yet the Defense Department continued development.[42] The V-22 program manager cited a rush to develop the technology as

33 The MITRE Corporation, "Perspectives on Research in Artificial Intelligence," 27, 55.

34 *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity*, Department of Defense, 2018, 8, 15, https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

35 There are some notable exceptions, such as the U.S. Navy submarine community, that have extremely low accident rates due to rigorous institutional procedures, such as the Navy's SUBSAFE program.

36 "Trends in Active-Duty Military Deaths Since 2006," Congressional Research Service, July 1, 2019, updated May 17, 2021, https://fas.org/sgp/crs/natsec/IF10899.pdf.

37 Zachary Cohen, "With 16 Service Members Killed in Air Crashes, Top Lawmaker Says 'Readiness of the Military Is at a Crisis Point," *CNN*, April 7, 2018, https://www.cnn.com/2018/04/07/politics/us-military-aviation-deaths-trump-national-guard/index.html.

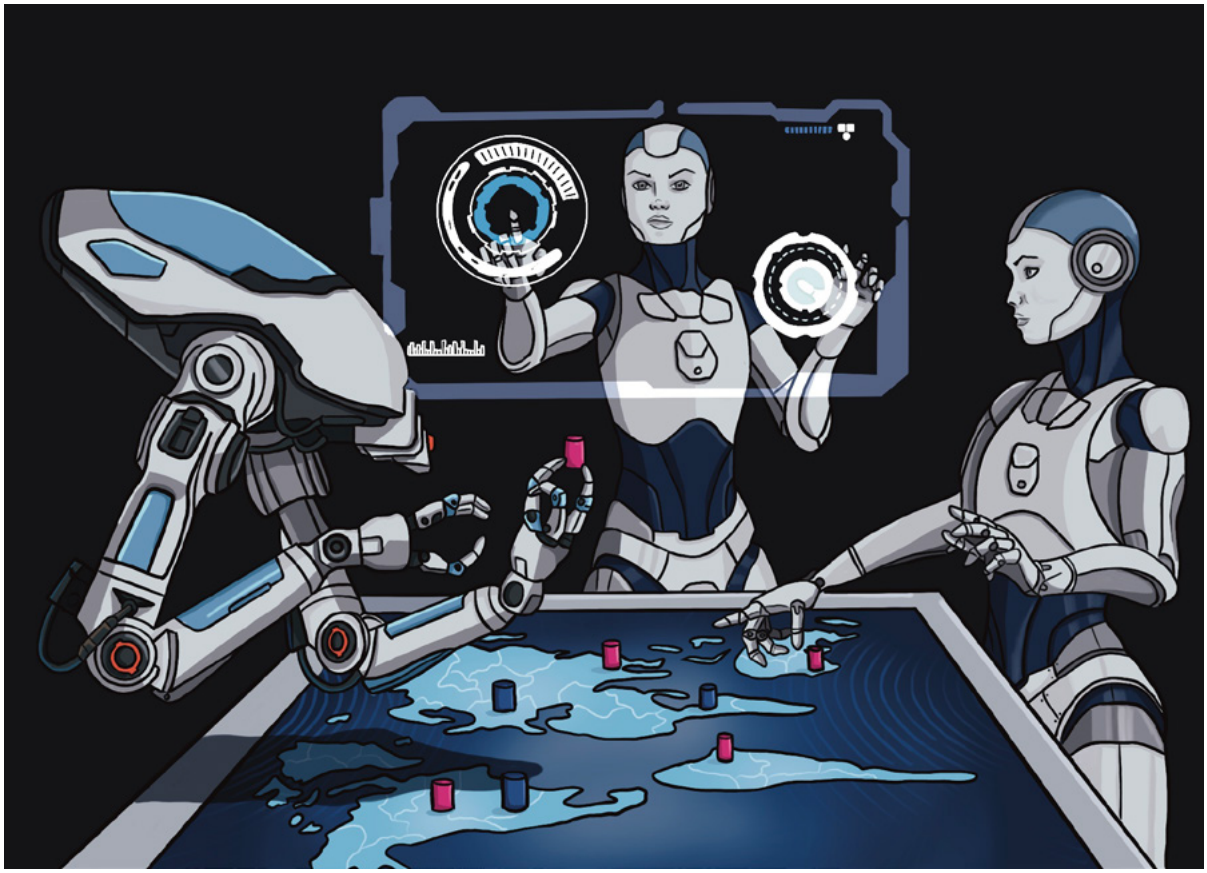
38 Erin Patterson, "Ship Collisions: Address the Underlying Causes, Including Culture," U.S. Naval Institute, August 2017, https://www.usni.org/magazines/proceedings/2017/august/ship-collisions-address-underlying-causes-including-culture.

39 "Buchanan Calls for Congressional Hearing on Military Training Accidents," Office of Representative Vern Buchanan, press release, Aug. 17, 2020, https://buchanan.house.gov/media-center/press-releases/buchanan-calls-congressional-hearing-military-training-accidents.

40 William M. Arkin and Joshua Handler, "Naval Accidents 1945–1988," Greenpeace Institute for Policy Studies, Neptune Paper No. 3, June 1989, https://fas.org/wp-content/uploads/2014/05/NavalAccidents1945-1988.pdf.

41 "Major Russian Submarine Accidents Since 2000," *Radio Free Europe/Radio Liberty*, July 2, 2019, https://www.rferl.org/a/major-russian-submarine-accidents-since-2000/30033592.html; and Peter Suciu, "Steel Tomb: The Worst Russian Submarine Disasters of All Time," *National Interest*, May 12, 2020, https://nationalinterest.org/blog/buzz/steel-tomb-worst-russian-submarine-disasters-all-time-153216.

42 Jeremiah Gertler, *V-22 Osprey Tilt-Rotor Aircraft: Background and Issues for Congress*, Congressional Research Service, March 2011, https://fas.org/sgp/crs/weapons/RL31384.pdf.



a factor in the accidents, stating, "Meeting a funding deadline was more important than making sure we'd done all the testing we could."[43] Taking shortcuts on testing in particular appears to have been a factor in at least one fatal crash. According to a Government Accountability Office investigation of the V-22 program, "schedule pressures" led the program to conduct only 33 of 103 planned tests of an aerodynamic phenomenon called a "vortex ring state,"[44] a phenomenon that later caused an April 2000 crash that killed 19 servicemembers.[45]

Absent competitive dynamics, militaries may be able to manage the challenges of fielding safe AI systems to a more-or-less satisfactory degree, albeit with some risk of an accident occurring. However, out of a desire to field AI capabilities ahead of competitors, militaries may be more willing to accept risk than they might otherwise be and to field systems that are prone to mishaps.[46] Similar competitive dynamics may have played a role in accidents with self-driving cars and commercial airline autopilot technology, as companies rushed to beat others to market.[47] These dynamics, while not an arms race, could lead militaries to engage in a "race to the bottom" on safety. This risk could become particularly acute in wartime.

Managing these risks is challenging because assessing them can be difficult, especially when it comes to new technologies. Accident rates may be well-known for mature technologies, but they are unknown for technologies still in development. In the case of the V-22 Osprey development, for example, it is not as though the Defense Department knew that developing it would lead to multiple crashes and 30 fatalities but decided that achieving the capability was worth the cost. Engineers, test-

43 Ron Berler, "Saving the Pentagon's Killer Chopper-Plane," *Wired,* July 1, 2005, https://www.wired.com/2005/07/osprey/.

44 Berler, "Saving the Pentagon's Killer Chopper-Plane"; and Mary Pat Flaherty and Thomas E. Ricks, "Key Tests Omitted on The Osprey," *Washington Post,* Feb. 19, 2001, https://www.washingtonpost.com/archive/politics/2001/02/19/key-tests-omitted-on-the-osprey/a657accd-d14c-4235-a781-6e05e12b25fd/.

45 Berler, "Saving the Pentagon's Killer Chopper-Plane."

46 Richard Danzig, "Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority," Center for a New American Security, May 2018, https://www.cnas.org/publications/reports/technology-roulette.

47 Charles Duhigg, "Did Uber Steal Google's Intellectual Property?" *The New Yorker,* Oct. 15, 2018, https://www.newyorker.com/magazine/2018/10/22/did-uber-steal-googles-intellectual-property; and David Gelles, et al., "Boeing Was 'Go, Go, Go' to Beat Airbus With the 737 Max," *New York Times,* March 23, 2019, https://www.nytimes.com/2019/03/23/business/boeing-737-max-crash.html.

ers, and program managers are flying in the dark when it comes to new technologies — that is, after all, the point of testing new systems. The concern is not only that organizations may take measured risks to field new capabilities, but also that institutional and bureaucratic imperatives may lead organizations to distort their own perceptions of risk, further contributing to accidents. This sociological phenomenon has been cited as a cause in the 1986 Space Shuttle *Challenger* explosion, for example.[48]

That militaries may incur risks from moving too quickly in adopting new technology is counter to the common caricature of military culture as conservative, hidebound, and resistant to innovation. While this caricature is not entirely fair — militaries do innovate even in peacetime[49] — the lack of direct, observational feedback on performance in a realistic competitive environment, akin to marketplace dynamics for commercial companies, can mean that militaries are often slow to adapt to changing circumstances. A variety of factors can affect military adoption of new technologies,[50] and adoption rates can vary considerably depending on the technology, state, and military community. Across a range of contemporary technologies militaries lag the private sector. For example, modern-day militaries are behind the private sector in adopting information technology, human performance optimizing technologies, and personnel best practices. With an increasing amount of technological innovation occurring outside of the defense sector, this lag is likely to continue.[51] Yet, the key risk factor for military AI systems is not the timing of when militaries begin the process of adoption, but the taking of shortcuts in safety to accelerate fielding new AI capabilities.

Adopting technology is a multi-stage process, involving research and development, experimentation, prototyping, technology maturation, production, testing, and fielding. It is possible for militaries to move slowly in one stage and quickly (or via shortcuts) in others. While there are many areas in which the U.S. military's adoption of AI, autonomy, robotics, and uninhabited vehicles is moving slowly due to a variety of bureaucratic obstacles, it is also possible that the United States could rush parts of the adoption process and end up with immature technology in production or even in the field. This mixed dynamic, of moving slowly in some aspects of technology development and taking shortcuts in others, has been present in other defense programs. The F-35 fighter jet went into production before the first test flight, a decision that the Defense Department's top acquisition official, Frank Kendall, later characterized as "acquisition malpractice."[52] Yet, the whole acquisition program took 25 years from its initial conception to its first operational deployment.[53] The F-35 is still not in full rate production, 28 years after it was initially conceived of.[54] The F-35 program moved too quickly in some areas, introducing unnecessary risk, even while it was overall encumbered by the laborious pace typical of major defense acquisition programs. Militaries' sluggish bureaucracy is, therefore, no defense against shoddy testing and premature fielding.

Costly signals, such as investing in AI safety research or TEVV processes and infrastructure, may be even more effective in demonstrating to other nations that a state values fielding safe AI systems that operate under effective human control.

48 Diane Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA* (Chicago: University of Chicago Press, 1996, 2016).

49 Stephen Peter Rosen, *Winning the Next War: Innovation and the Modern Military* (Ithaca, NY: Cornell University Press, 1994).

50 Michael C. Horowitz, *The Diffusion of Military Power: Causes and Consequences for International Politics* (Princeton, NJ: Princeton University Press, 2010).

51 *The Global Research and Development Landscape and Implications for the Department of Defense*, Congressional Research Service, Nov. 8, 2018, https://fas.org/sgp/crs/natsec/R45403.pdf.

52 Colin Clark, "F-35 Production Move Was 'Acquisition Malpractice': Top DoD Buyer," *Breaking Defense*, Feb. 6, 2012, https://breakingdefense.com/2012/02/f-35-production-was-acquisition-malpractice-top-dod-weapons-b/.

53 *F-35 Joint Strike Fighter (JSF) Program*, Congressional Research Service, May 27, 2020, https://fas.org/sgp/crs/weapons/RL30563.pdf; The Joint Advanced Strike Technology (JAST) program, which later became the Joint Strike Fighter program, was created in 1993. "JAST: History," JSF.mil, archived as of April 1, 2006, https://web.archive.org/web/20190715052740/http://www.jsf.mil/history/his_jast.htm; and "Lockheed F-35 Jet Used by U.S. in Combat for First Time: Official," *Reuters*, Sept. 27, 2018, https://www.reuters.com/article/us-usa-pentagon-f35/lockheed-f-35-jet-used-by-u-s-in-combat-for-first-time-official-idUSKCN1M72BT.

54 John A. Tirpak, "F-35 Full-Rate Still Months Away, But Won't Signal Production Surge," *Air Force Magazine*, March 17, 2021, https://www.airforcemag.com/f-35-full-rate-still-months-away-but-wont-signal-production-surge/.

Avoiding the Harmful Risks of AI Security Competition

What can states do to avoid a race to the bottom on safety or an acceleration of the tempo of war beyond human control? In both instances, there are countervailing incentives that push against these trends. Militaries desire trusted systems on the battlefield and effective control over their own forces. There are several actions states can take to strengthen these incentives toward ensuring robust, secure, and controllable AI systems in their own institutions, as well as those of other countries.

First, states should invest in adequate internal processes to test, evaluate, verify, and validate AI systems, in order to ensure that the systems they are fielding are robust and secure.[55] States should similarly strengthen their internal processes — doctrine, training, system design and testing, human-machine interfaces, etc. — to retain effective human control over combat operations.[56]

Second, states should take specific measures to encourage other states to do likewise in order to mitigate perverse incentives to cut corners on testing or cede human control to machines where it would otherwise not be preferable. Such actions could include voluntary transparency measures about TEVV processes, although there will no doubt be technical details that states are unwilling to share. States could also communicate the importance of AI safety and reliability and of maintaining human control over combat operations, both publicly and in international diplomatic channels such as the Convention on Certain Conventional Weapons.[57] For example, in 2020 the U.S. Department of Defense released a set of ethical principles for AI.[58] Costly signals, such as investing in AI safety research or TEVV processes and infrastructure, may be even more effective in demonstrating to other nations that a state values fielding safe AI systems that operate under effective human control. States should avoid messages that may incentivize other states to take shortcuts on these processes, such as claims of an "AI arms race."

Lastly, states should explore opportunities to take cooperative measures that might mitigate these risks. Getting adversaries to cooperate is inherently challenging, but states have succeeded in the past in regulating the conduct of war in a variety of ways to mitigate mutual harm. Joint declarations, codes of conduct, or confidence-building measures may help to reduce the greatest dangers of AI competition and encourage states to adopt AI responsibly.[59]

The United States has done more to date than any other nation to advance norms surrounding the responsible use of AI, although the Department of Defense could be more deliberate in its approach to addressing the risks of military AI competition. U.S. defense leaders have focused primarily on implementing and demonstrating AI applications in an effort to prove AI's value in military operations. This is understandable. The Department of Defense has many practical challenges to fielding AI systems even in relatively low-risk applications, including problems with data, computing infrastructure, contracting, and funding.[60] Nevertheless, it can and should do more to ensure that, as it competes in AI, it does so in a way that does not generate unnecessary risks or undermine international stability.

The most important step that defense leaders could take in the near term to mitigate the risks stemming from AI competition would be to implement the necessary internal processes to ensure adequate TEVV of AI systems. A 2019 congressionally mandated independent assessment of the Defense Department's AI efforts conducted by the RAND Corporation found that current TEVV processes were "nowhere close to ensuring the performance and safety of AI applications, particularly where

55 Flournoy, Haines, and Chefitz, "Building Trust through Testing."

56 "Final Report," National Security Commission on Artificial Intelligence, March 2021, https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf, 131–40.

57 National Security Commission on Artificial Intelligence, "Final Report," 99.

58 "DOD Adopts Ethical Principles for Artificial Intelligence," U.S. Department of Defense, press release, Feb. 24, 2020, https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

59 Andrew Imbrie and Elsa Kania, "AI Safety, Security, and Stability Among Great Powers: Options, Challenges, and Lessons Learned for Pragmatic Engagement," Center for Security and Emerging Technology, December 2019, https://cset.georgetown.edu/publication/ai-safety-security-and-stability-among-great-powers-options-challenges-and-lessons-learned-for-pragmatic-engagement/; Giacomo Persi Paoli, et al., "Modernizing Arms Control," United Nations Institute for Disarmament Research, Aug. 30, 2020, https://www.unidir.org/publication/modernizing-arms-control; and Michael C. Horowitz, Lauren Kahn, Casey Mahoney, "The Future of Military Applications of Artificial Intelligence: A Role for Confidence-Building Measures?" *Orbis* 64, no. 4 (2020): 528–43, https://doi.org/10.1016/j.orbis.2020.08.003; Michael Horowitz and Paul Scharre, "AI and International Stability: Risks and Confidence-Building Measures," Center for a New American Security, Jan. 12, 2021, https://www.cnas.org/publications/reports/ai-and-international-stability-risks-and-confidence-building-measures; and National Security Commission on Artificial Intelligence, "Final Report," 97–101.

60 Rachel S. Cohen, "Roper Argues for More Artificial Intelligence Collaboration, Funding," *Air Force Magazine*, Sept. 10, 2020, https://www.airforcemag.com/roper-argues-for-more-artificial-intelligence-collaboration-funding/; and National Security Commission on Artificial Intelligence, "Final Report," 59–74.

safety-critical systems are concerned," and issued recommendations for addressing this gap.[61] Similarly, a 2020 independent study led by Michèle Flournoy and Avril Haines identified a number of actionable steps that the department could take to improve its AI TEVV.[62] The National Security Commission on AI also concluded that "TEVV of traditional legacy systems is not sufficient" at providing adequate assurance for AI systems, and that "an entirely new type of TEVV will be needed."[63] The report issued a raft of recommendations to improve AI TEVV and establish "justified confidence in AI systems."[64]

The Defense Department should adopt these reports' recommendations to improve AI TEVV, a step that would not only bolster the safety of its AI systems but also their effectiveness. In addition to increasing resources and getting the attention of senior leadership, improving TEVV will require shifting how senior Defense Department leaders think about building robust, reliable, and effective AI systems. At times, senior defense leaders have characterized safety and ethics as an encumbrance the United States has to deal with that its adversaries do not.[65] While it is undoubtedly true that Russia and China are less concerned about ethics, safety, or international law than the United States, ensuring that military AI systems operate effectively and in a way that is consistent with human intent is a strength in the long run, even if more rigorous TEVV processes are needed in the near term to achieve that goal.

Another element of mitigating the risks of military AI competition is with regard to how states characterize AI. U.S. messaging has been consistent and strong on the need for the responsible, lawful, ethical, and safe use of AI.[66] In their messaging, however, U.S. policymakers have frequently refrained from highlighting the risks of military AI competition, such as those outlined in this article. At times, they have emphasized a desire for speed that could feed into security dilemma concerns about a race-to-fielding that could undermine safety. In his 2020 *Wired* article, Will Roper wrote: "Our nation must wake up *fast*. The only thing worse than fearing AI itself is fearing not having it."[67] While he called for using AI "safely and effectively," his overriding emphasis was for the Defense Department to move faster.[68]

Quite understandably, U.S. policymakers working to accelerate the adoption of AI in a slow-moving and sclerotic bureaucracy may be loath to portray the technology as immature, unready, or unreliable. Additionally, U.S. policymakers may fear that highlighting the risks of military AI could contribute to AI engineers balking at working with the military, even if such fears are unfounded.[69] Nevertheless, just as other emerging technologies such as computer networks opened up novel strategic challenges in the form of cyber operations, defense analysts should begin to think now about ways that AI may complicate international stability. A forthright acknowledgment of the risks of military AI competition is the first step toward a strategy of competing smartly while addressing those risks. Acknowledging the risks of military AI competition does not mean that the United States should refrain from adopting AI any more than acknowledging the stability risks of competing in space, cyberspace, or nuclear weapons necessitates unilateral disarmament in those spheres. America's response to these risks should not be to refrain from AI competition, but rather to shape the character of the competition so that states are, at a minimum, mindful of these risks.

The National Security Commission on AI has demonstrated what such an approach might look like in practice. The commission's 700-plus page report issued sweeping recommendations for enhancing U.S. competitiveness in AI and military adoption, but also dedicated an entire chapter to "Autonomous Weapon Systems and Risks Associated with AI-Enabled Warfare."[70] With regards to safety concerns, the report acknowledged:

61 Tarraf et. al., "The Department of Defense Posture for Artificial Intelligence," xiii, xv.

62 Flournoy, Haines, and Chefitz, "Building Trust through Testing."

63 National Security Commission on Artificial Intelligence, "Final Report," 137.

64 National Security Commission on Artificial Intelligence, "Final Report," 131–40.

65 Roper, "There's No Turning Back on AI in the Military."

66 Department of Defense, "Summary of the 2018 Department of Defense Artificial Intelligence Strategy"; Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence"; "Responsible AI Champions Pilot," Joint Artificial Intelligence Center, Department of Defense, press release, accessed Nov. 12, 2020, https://www.ai.mil/docs/08_21_20_responsible_ai_champions_pilot.pdf; and "AI Partnership for Defense," Joint Artificial Intelligence Center, Department of Defense, press release, accessed Nov. 12, 2020, https://www.ai.mil/docs/AI_PfD_Joint_Statement_09_16_20.pdf.

67 Roper, "There's No Turning Back on AI in the Military."

68 Roper, "There's No Turning Back on AI in the Military."

69 Catherine Aiken, Rebecca Kagan, and Michael Page, "'Cool Projects' or 'Expanding the Efficiency of the Murderous American War Machine?'" Center for Security and Emerging Technology, November 2020, https://cset.georgetown.edu/research/cool-projects-or-expanding-the-efficiency-of-the-murderous-american-war-machine/.

70 National Security Commission on Artificial Intelligence, "Final Report," 89–106.

> Russia and China are likely to field AI-enabled systems that have undergone less rigorous TEVV than comparable U.S. systems and may be unsafe or unreliable ... The United States should ... highlight how deploying unsafe systems could risk inadvertent conflict escalation [and] emphasize the need to conduct rigorous TEVV.[71]

The commission issued a number of recommendations to mitigate the risks of AI competition, including improving Defense Department TEVV processes and working with allies to develop "international standards of practice for the development, testing, and use of AI-enabled and autonomous weapon systems" to reduce the risk of accidents.[72]

Recently, the Defense Department has taken positive steps toward emphasizing AI safety and improving TEVV processes. In May 2021, Deputy Secretary of Defense Kathleen Hicks issued a memorandum on implementing "Responsible AI." The memo launched a series of internal bureaucratic structures, including "establishing a test and evaluation and verification and validation framework that integrates real-time monitoring, algorithm confidence metrics, and user feedback to ensure trusted and trustworthy AI capabilities."[73] Additionally, in public comments Hicks emphasized the importance of AI "safety."[74] These are important and valuable steps toward establishing the necessary bureaucratic processes to ensure U.S. military AI systems are robust and reliable as well as setting a constructive tone publicly. The United States has been active in promulgating norms about the responsible use of military AI. A deliberate approach toward acknowledging and mitigating the risks of AI competition need not come at the expense of adopting AI to improve military effectiveness.

Ideally, a frank assessment of the risks of AI competition and U.S. transparency about measures it is taking to mitigate these risks would open the door to cooperative measures among competitors. There may be a variety of confidence-building measures that states could adopt to reduce the risks of AI competition.[75] "Track II" dialogues among academic experts to better understand these risks and potential cooperative measures are already underway. Future direct government-to-government dialogues could explore whether there is opportunity for common ground. Cooperative measures to reduce risk will depend on other states such as Russia and China engaging in good faith. However, there is no guarantee that they will do so. What the United States can do is improve its own internal processes for AI TEVV and for ensuring human responsibility. The United States should also publicly articulate why it would be in other states' best interests to cooperate to avoid some of these mutual risks. Even as the United States adopts AI to improve its national defense, it should take measures — and incentivize others to do so as well — to ensure military AI systems are safe and that warfare remains under effective human control.

Paul Scharre *is the vice president and director of studies at the Center for a New American Security and is the author of* Army of None: Autonomous Weapons and the Future of War. *This article was made possible, in part, due to grants from Open Philanthropy and Carnegie Corporation of New York. Portions of this article are adapted from the author's doctoral thesis, "Autonomous Weapons and Stability."*

Image: U.S. Navy/Peggy Frierson

71 National Security Commission on Artificial Intelligence, "Final Report," 99.

72 National Security Commission on Artificial Intelligence, "Final Report," 97–101.

73 "Memo: Implementing Responsible Artificial Intelligence in the Department of Defense," Deputy Secretary of Defense, May 26, 2021, https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/IMPLEMENTING-RESPONSIBLE-ARTIFICIAL-INTELLIGENCE-IN-THE-DEPARTMENT-OF-DEFENSE.PDF.

74 Patrick Tucker, "US Needs to Defend Its Artificial Intelligence Better, Says Pentagon No. 2," *Defense One*, June 22, 2021, https://www.defenseone.com/technology/2021/06/us-needs-defend-its-artificial-intelligence-better-says-pentagon-no-2/174876/.

75 For more details on potential AI confidence-building measures, see Horowitz and Scharre, "AI and International Stability," 12–21.