U.S.-China Competition and Military AI
How Washington Can Manage Strategic Risks amid Rivalry with Beijing

Jacob Stokes and Alexander Sullivan with Noah Greene
About the Author

Jacob Stokes is a senior fellow in the Indo-Pacific Security Program at CNAS, where his work focuses on U.S.-China relations, Chinese foreign policy, East Asian security affairs, and great power competition. He previously served in the White House on the national security staff of then–Vice President Joe Biden. He also has worked as a professional staff member for the U.S.-China Economic and Security Review Commission, an advisor for U.S. Senator Amy Klobuchar, and a senior analyst in the China program at the U.S. Institute of Peace. He is a term member of the Council on Foreign Relations.

Alexander Sullivan is an Adjunct Fellow in the Indo-Pacific Security Program at CNAS, where he focuses on U.S.-China relations, Chinese foreign policy, and the nexus of economics and security.

Noah Greene is the Project Assistant for the AI Safety & Stability Project at CNAS. His portfolio includes research on the military applications of artificial intelligence in the United States, Russia, China, and Europe.

About the Indo-Pacific Security Program

The CNAS Indo-Pacific Security Program addresses opportunities and challenges for the United States in the region, with a growing focus on issues that originate in the Indo-Pacific but have global implications. It draws on a team with deep government and nongovernment expertise in regional studies, U.S. foreign and defense policy, and international security. The Indo-Pacific Security Program analyzes trends and generates practical and creative policy solutions around five main research priorities: U.S.-China strategic competition, India’s growing role in the Indo-Pacific, American alliances and partnerships, the North Korea threat, and challenges in South and Central Asia.

About the Artificial Intelligence Safety & Stability Project

The CNAS AI Safety & Stability Project is a multiyear, multiprogram effort that addresses the established and emerging risks associated with artificial intelligence. Our work is focused on anticipating and mitigating catastrophic AI failures; improving the U.S. Department of Defense’s processes for AI testing and evaluation; understanding and shaping opportunities for compute governance; understanding Chinese decision-making on AI and stability; and understanding Russian decision-making on AI and stability.

Acknowledgments

The authors are grateful to the many officials and experts—both in the United States and elsewhere—who shared their perspectives over the course of this project. They are particularly indebted to Joel Wuthnow, Kevin Pollpeter, J. Michael Dahm, Jacquelyn Schneider, Adam Segal, and James Acton for providing insightful presentations to the working groups for this project. Thanks are also due to Vincent Boulanin, Paul Scharre, and Lisa Curtis for their expert reviews of the manuscript, as well as to Maura McCarthy for her sharp editing. Finally, this paper would not have been possible without assistance from a host of CNAS colleagues, including Rebecca Wittner, Melody Cook, Rin Rothback, Allison Francis, Jake Penders, Emma Swislow, and Lesly Melendez. The views presented here do not represent those of CNAS or any other organization, and the authors are solely responsible for any errors in fact, analysis, or omission. This report was made possible with the generous support of Open Philanthropy.

As a research and policy institution committed to the highest standards of organizational, intellectual, and personal integrity, CNAS maintains strict intellectual independence and sole editorial direction and control over its ideas, projects, publications, events, and other research activities. CNAS does not take institutional positions on policy issues and the content of CNAS publications reflects the views of their authors alone. In keeping with its mission and values, CNAS does not engage in lobbying activity and complies fully with all applicable federal, state, and local laws. CNAS will not engage in any representational activities or advocacy on behalf of any entities or interests and, to the extent that the Center accepts funding from non-U.S. sources, its activities will be limited to bona fide scholastic, academic, and research-related activities, consistent with applicable federal law. The Center publicly acknowledges on its website annually all donors who contribute.
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Executive Summary</td>
</tr>
<tr>
<td>02</td>
<td>Introduction</td>
</tr>
<tr>
<td>02</td>
<td>Beijing’s AI Plans and AI’s Role in China’s Military Modernization</td>
</tr>
<tr>
<td>06</td>
<td>Strategic Risk Pathways Military AI Could Create or Exacerbate in U.S.-China Relations</td>
</tr>
<tr>
<td>11</td>
<td>Options for Managing Strategic Risks from Military AI</td>
</tr>
<tr>
<td>14</td>
<td>Recommendations for Policymakers</td>
</tr>
<tr>
<td>17</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>
Executive Summary

Two tectonic trends in the international security environment appear to be on a collision course. The first trend is the intensifying geopolitical rivalry between the United States and the People’s Republic of China (PRC or China). The second trend is the rapid development of artificial intelligence (AI) technologies, including for military applications. This report explores how the United States can manage strategic risks—defined as increased risks of armed conflict or the threat of nuclear war—that could be created or exacerbated by military AI in its relationship with China.

It begins by providing an overview of China’s views on and policies toward AI. Beijing sees AI playing roles in both its civilian economy and the modernization of its military, the People’s Liberation Army (PLA). At home, Chinese leaders want to leverage AI to boost growth and innovation, address economic and social challenges, and secure the Chinese Communist Party’s (CCP) domestic rule.

AI also plays a key role in China’s military ambitions, especially its goal to become a “world-class military” by midcentury, in part through the “intelligentization” of its forces. Intelligentization relies on integrating AI and other emerging technologies into the joint force with the goal of gaining an edge on the United States. China argues that its governance model, including its military-civil fusion policy, gives Beijing a competitive advantage over Washington. Realization of that vision, however, remains uncertain and will require China to overcome external and internal obstacles.

Next, the report articulates five categories of what the authors call pathways, or causal links, through which applications of military AI could undermine stability and increase strategic risk between Washington and Beijing. The first is individual improvements in capabilities that combine to give China a military edge. The second is AI’s effects on the decision-making and information domain. The third is uncrewed autonomous systems. The fourth is intelligence, surveillance, and reconnaissance. And the fifth is command, control, and communications.

The report’s discussion of each pathway provides more details about the intricacies of how they might function. Taken together, the emergence of military AI will likely deepen U.S.-China rivalry and increase strategic risks. Policymakers can draw on three categories of response options. These include limiting China’s military AI while advancing the United States’ own capabilities; engaging in unilateral responsible management of military AI; and pursuing bilateral and multilateral diplomacy to reduce strategic risks. Washington has already taken actions across all three categories, and foreign capitals have likewise sought to shape the military AI environment. To build on the actions taken to date, U.S. policymakers should:

- Take bold action to constrain China’s progress in AI for military and repressive purposes, but do so in a narrow way that avoids self-defeating steps;
- Build U.S. military AI capabilities to stay on the cutting edge;
- Develop, promulgate, and implement norms and best practices on responsible military AI;
- Proactively engage with like-minded allies and partners and in multilateral institutions on military AI issues;
- Negotiate risk reduction and confidence-building measures with China related to military AI;
- Continue to pursue universal U.S.-China risk reduction and crisis management mechanisms despite persistent challenges;
- Make military AI a fundamental pillar of diplomacy with China related to nuclear weapons and strategic stability;
- Take steps to reduce strategic risks not directly caused by military AI but potentially worsened by the inherent speed and unpredictability of military AI;
- Prioritize intelligence-gathering and analysis on, and net assessment of, China’s military AI capabilities.
Beijing’s AI Plans and AI’s Role in China’s Military Modernization

Both public and private actors within the PRC view artificial intelligence as crucial to China’s future. For the party-state, AI’s importance goes beyond any contribution to military or political power, though those are of course important benefits. It regards AI as key to enhancing economic growth and social welfare—in short, as critical to the country’s future in every respect.

China’s top leaders have repeatedly stated their commitment to making the country into, as Chinese Communist Party (CCP) General Secretary Xi Jinping put it in his work report to the 19th Party Congress in October 2017, a “science and technology great power.” At the 20th Party Congress in October 2022, Xi stated that “innovation will remain at the heart of China’s modernization drive”; that by 2035, China would “join the ranks of the world’s most innovative countries, with great self-reliance and strength in science and technology”; and that it should “resolutely win the battle over key and core technologies.” This commitment by Xi was reaffirmed in an April meeting of the CCP Politburo in which he called for the “development of general artificial intelligence,” often referred to in the West as artificial general intelligence (AGI) and defined as a form of AI capable of broad problem-solving across different environments. The definition of AGI is somewhat vague, though, and often changes with new advancements in the technology.
As the 2017 “New Generation Artificial Intelligence Development Plan” (AIDP) put it, “AI has become a new engine of economic development.”6 Given that data is one of the key inputs to AI development, China’s AI goals were a key motivation to create a National Data Administration as part of the overall reorganization of the State Council announced at the 2023 “Two Sessions” meetings.9 China’s leaders believe world-leading AI capabilities are necessary for their country to continue climbing the global value chain and to move beyond its legacy development model—which has delivered rapid gains in gross domestic product (GDP) but produced geographic imbalances, environmental costs, and industrial overcapacity—to “high-quality” and sustainable growth.

Chinese planners look to applications of AI in specific fields to address myriad economic and social challenges.10 They expect AI and related technologies like robotics to increase industrial productivity, offsetting the shrinking of China’s working-age population while also better harnessing the increased human capital of younger generations who are better educated than their predecessors. “Smart” agriculture, cities, transit, and logistics will enable the continued urbanization and transition to domestic consumption that underpin China’s medium- and long-term economic plans. Leaders in Beijing also view AI as the next step in a long-standing drive to improve public service provision through digitization of government functions. AI-enabled advances in medicine promise to deliver better health care outcomes at reduced cost and help alleviate the burdens of caring for an aging populace. More efficient and dynamic planning and use of resources are key to accomplishing all these objectives while meeting China’s climate and environmental protection targets.

The CCP values AI not only for economics and welfare, but also for securing its domestic rule and projecting its power internationally. At a basic level, sustaining China’s decades-long buildup of military and internal security forces depends upon resources that can only be delivered from continued economic growth. Moreover, AI is knit into the CCP’s intricate system of social control. While narratives of an AI-driven, all-seeing “social credit system” describe Beijing’s ambition more than the current technological reality,11 the use of automated content moderation is already a crucial part of China’s vast system of internet censorship.12 Facial and other biometrics recognition, computer vision, and other AI technologies are key tools in the genocide against Uyghurs in Xinjiang. Those techniques are increasingly employed to monitor and control other minority groups and the broader population, too.13

Finally, while the CCP aims to direct and coordinate the development and use of AI throughout Chinese society, the party recognizes that it cannot do so alone. Private actors in business, academia, medicine, and civil society will be the primary ones figuring out how to use these tools to push the boundaries in their fields. For its part, the party aims to create public goods by investing in both basic and applied AI research, while yoking AI developments to its own goals.14 Mechanisms for this approach include the creation of a “national team” of large firms chosen to lead in specific fields, the use of industry alliances between firms and various levels of government, and the decentralized local experimentation that has long characterized Chinese economic policymaking.15 In addition, China’s system of Military-Civil Fusion—although its scope remains ambiguous—seeks to appropriate select private technological advancements, including some developed in cooperation with international research partners, to augment the PLA’s capabilities.16

China’s Unique Model for AI Development and Deployment

The Chinese government’s ability to marshal and coordinate resources to push artificial intelligence development has led many commentators to predict that it will inevitably “win the race” for AI.17 In addition, there is some evidence that Chinese society overall is more receptive to these technologies: a 2022 Ipsos poll found that supermajorities of Chinese respondents were highly aware of AI progress and felt that it would change their lives for the better. Less than half of U.S. respondents felt similarly.18
However, a number of institutional factors may shape the extent to which China can become the world’s leader on AI. First, like others around the globe, China’s government has sought to articulate ethical limits guiding the type of AI that is developed. In 2019, for example, the Ministry of Science and Technology published “governance principles” for “responsible AI,” which urged that AI development should be prosocial, fair, inclusive, privacy-respecting, safe, controllable, open, and agile. The AIDP anticipates that such regulations will be expanded and developed into laws throughout the plan’s 2030 timeline.

Second, China has placed limits on industry’s use of data, which is a critical input for AI development. The government expects private companies—especially those on its “national team”—to lead the way in key areas of AI development. At the same time, responding to public outcry over repeated breaches of privacy, China has issued several binding regulations to limit those companies’ ability to exploit private customer data. The 2021 Personal Information Protection Law and Data Security Law mandates that data handlers obtain consumers’ informed consent about data collection and use, create suitable procedures for protecting data against unauthorized disclosure, and establish nontrivial penalties for enterprises that violate these rules.

As part of the overall reorganization of the State Council announced at the 2023 “Two Sessions” meetings, China created a new National Data Administration that will, among other things, both enable and constrain industry’s use of data for AI development. However, these laws explicitly establish that no privacy restrictions will apply to state security organs, which continue to conduct increasingly AI-enabled mass surveillance on the population. In general, the Chinese approach has been the inverse of the United States’. In America, there are legal and institutional limits on what information the state can collect and maintain on citizens but almost no constraints on private enterprises’ data collection and use outside of select industries such as finance or health care.

Finally, the PRC government is moving to limit the ways in which businesses and other actors can use AI algorithms within China. In March 2022, new regulations on internet services’ algorithmic recommender systems went into effect, administered by the Cyberspace Administration of China (CAC). These regulations ostensibly seek to prevent monopolistic behavior, such as anticompetitive walled gardens or excessive price discrimination, and prevent the spread of fake news—again, a contrast to the laissez-faire approach of the United States. They also created an “algorithm registry” intended to give the government greater visibility into and oversight of businesses’ AI-powered services and the data used to train them.

In the wake of ChatGPT’s explosion, China also aims to regulate generative AI. CAC issued draft regulations in April 2023 intended to control both the systems’ training data and their output. Not only must training data comply with the aforementioned privacy laws, but it also must respect copyright and be “accurate.” Moreover, the rules required that outputs be “true and accurate”—a response to the danger that inherently unpredictable AI systems may spit out content the CCP deems politically subversive. Interim regulations released in July 2023 softened the rules, though, by exempting non-public-facing R&D activities and jettisoning the false content liability so that it requires only “effective measures” to increase accuracy. It is unclear how much the need for political control will undermine China’s ability to fully harness generative AI technology.

It remains to be seen whether China’s domineering regulatory stance leads in practice to reduced societal risks from AI. Both individual officials and the party at large have the incentives to downplay bad news and the tools to do so. These information problems can mask ongoing issues, allowing risks to build up out of sight. This dynamic could impair implementation of China’s AI regulations or hinder efforts to update them as conditions evolve.

In sum, as it pursues diverse breakthroughs in AI, China will have to continually manage tradeoffs between the openness that fosters innovation, the need to regulate business excesses, and its own desire to maintain tight political control over every part of Chinese society. Given these trends, U.S. policymakers and experts will need to eschew preconceived notions about China’s strengths and weaknesses and continually update their assessments as this dynamic field progresses.

**AI and China’s Military Modernization**

China also sees AI playing a critical role in advancing its military power. Xi has set ambitious goals for the PLA to “basically complete” its modernization by 2035 and transform into a “world-class” military by the middle of the century. In March 2023, Xi called on the PLA to “raise the presence of combat forces in new domains and of new qualities.” As part of those goals, Xi wants the PLA to continue to move through stages of military-technological development, from mechanization to informatization and ultimately intelligence. Broadly, mechanization refers to fielding modern platforms and equipment; informatization...
refers to linking those systems to networks such as GPS; and intelligentization refers to integrating artificial intelligence, quantum computing, big data, and other emerging technologies into the joint force. In 2020, China set a new goal to “accelerate the integrated development of mechanization, informatization, and intelligentization” by 2027. In other words, Beijing aims to make progress on all three stages simultaneously rather than sequentially.

Beijing sees progressing through these stages as necessary to keep up with changes in the technological character of warfare in the 21st century. Chinese scholars speak about the ongoing revolution in military affairs as one of weapons “systems confrontation” that requires “systems destruction warfare” to win. To compete in this emerging era of conflict, the PLA is developing an overarching concept it calls “multidomain precision warfare.” In layman’s terms, this concept posits that the very networking that gives the U.S. military its power creates vulnerabilities that can be exploited. Thus, rather than needing to destroy U.S. enemy forces directly—ship-to-ship or tank-to-tank—China can attack the weak points that link U.S. systems and domains together and thereby neutralize or overwhelm U.S. advantages. Those weak points can include internet, satellite, or electromagnetic communications links as well as logistical supply systems. AI is a critical part of this strategy because, in the dynamic environment of an actual conflict, identifying and targeting U.S. vulnerabilities will require sensing, relaying, and processing vast amounts of information at a speed only computers can match.

At the level of capabilities, the specific roles for military AI within China’s overall program of military modernization are still coming into focus. Researchers at the Center for Security and Emerging Technology analyzed 343 PLA equipment contracts and found seven areas of interest for current AI investments: (1) intelligent and autonomous vehicles; (2) intelligence, surveillance, and reconnaissance; (3) predictive maintenance and logistics; (4) information and electronic warfare; (5) simulation and training; (6) command and control; and (7) automated target recognition. Those categories are illustrative but not necessarily exhaustive. Moreover, AI is a general-purpose technology like electricity or railroads, so analysts cannot yet know all of its potential uses or implications. In the near- and mid-term, most of the changes AI will usher in will be incremental and narrow. But in the mid- to long-term, some could be revolutionary and general.

In the near- and mid-term, most of the changes AI will usher in will be incremental and narrow. But in the mid-to-long-term, some could be revolutionary and general. Potentially crowding out investments in next-generation AI-enabled capabilities. Should China’s domestic economy face sustained headwinds, fewer resources might be available for advancing AI within the PLA.

Furthermore, the technology itself might prove difficult to master even with abundant resources. Technology controls imposed by the United States and its allies could hamper Beijing’s ability to develop and operate AI-enabled systems at scale (see additional details below). In addition, China could simply lack the capacity to innovate at the bleeding edge of military technology. In earlier stages of its military modernization, China could imitate the United States, Russia, and other advanced militaries. By contrast, intelligentization requires pioneering totally new military technologies and operational concepts for how to use them.

Impediments related to personnel, bureaucratic structure, or political control could further constrain the PLA’s AI ambitions. These include a lack of skilled personnel needed to operate AI systems and stove-piped military bureaucracies. The PLA’s Strategic Support Force (SSF)—a stand-alone military service created in 2015 to focus on space, cyber, and electromagnetic warfare—appears to control the lion’s share of AI development resources and authority within the PLA.
While the SSF may have been created in part to enable jointness through advanced networking and now AI, it may be loath to relinquish control of its creations to the rest of the PLA, or other services might resist relying on capabilities run by the SSF.58 Finally, the CCP values control above other aims. The dictum that “the party controls the gun”—first stated by Mao Zedong and reaffirmed by Xi—and the prominent role of political commissars in the PLA reflect that fact.39 Even for expert researchers, today’s state-of-the-art AI models present challenges for predictability, “explainability,” and transparency. This opacity could make commanders reluctant to trust it for fear that they do not control its actions.40 Alternatively, however, Chinese leaders might be more willing to place trust in programmable machines over people. Given these contradictory impulses, it is not yet clear to what extent China’s military leadership and operational-level commanders will embrace or avoid AI.

**Strategic Risk Pathways Military AI Could Create or Exacerbate in U.S.-China Relations**

China, along with the United States, is moving quickly to integrate AI into its military. Many analysts have observed that this could feed instability in the already-tense security competition between Washington and Beijing.41 But exactly how emerging military AI capabilities could increase strategic risks is not always clear. This section addresses that gap by detailing five categories of pathways through which applications of military AI could undermine stability and increase strategic risk between the United States and China. Some of the military AI applications envisioned below are already available, while others are still emerging or only at the conceptual stage. Moreover, some strategic risks derive from AI working as intended. Others derive from AI acting in ways that are unintended.

**Pathway 1: Individual Capability Improvements That Combine to Give China a Military Edge**

Perhaps the most likely source of U.S.-China strategic risks stemming from military applications of AI will be one that is difficult to measure precisely and cannot be solely attributable to AI: the overall military balance. Many of the most practical uses for military AI in the near term will be for purposes that are relatively mundane but could help the PLA use resources more efficiently and therefore generate more military capabilities per yuan or dollar spent.42 These include helping to improve processes for maintenance, logistics, training, and decision-support.43 Such “back office” tasks rarely receive the sustained attention devoted to “tip of the spear” capabilities that appear on the front lines of combat. However, the strength of modern militaries depends as much on their enabling bureaucracies as their frontline troops and weapons.

In addition, some emerging military AI systems will improve the PLA’s combat capabilities. Initially, those improvements are likely to be evolutionary rather than revolutionary. Consider the air domain, to name just one of many examples. “Loyal wingman”-type systems where unmanned aircraft fly with manned aircraft could improve on what human pilots could do on their own.44 But fully autonomous and uncrewed air systems—capable of greater persistence, maneuverability, and other attributes due to their lack of human bodily limitations—will likely be necessary for a complete paradigm shift in air combat operations.

A similar story is playing out across nearly every aspect of military affairs. If, when improvements in military AI across every area are added together, the U.S.-China military balance turns in Beijing’s favor, then the risks of conflict could rise. China’s assertiveness on its periphery has grown in line with the PLA’s rapid modernization campaign since the 1990s, which has transformed the PLA into the world’s second-most-powerful military. In East Asia, U.S. military dominance, in conjunction with a network of American security allies and multilateral organizations, has underwritten a “long peace” free from major-power conflict in Asia since at least 1979.45 China’s military rise and attendant coercive behavior threaten to undermine that peace, especially if Beijing’s combat capabilities in the region eclipse Washington’s. As the U.S. National Security Strategy states, China is “the only competitor with both the intent to reshape the international order and, increasingly, the economic, diplomatic, military, and technological power to do it.”46

Of course, U.S. and allied actions will also shape the regional military balance, which is dynamic rather than static. Washington and its allies are trying to leverage AI to bolster their military capabilities, just as Beijing is. The unfortunate but inescapable reality is that East Asia is seeing a military-technological arms race, of which military AI is a part. And one way that China could eventually outgun the United States and its allies is to develop and field technologies such as military AI that would give Beijing a substantial advantage.
timelines could be to accelerate the tempo of crises and provide leaders with less decision time overall. If computational tools can perform some of these tasks faster than humans, states could fall into hasty decision-making just to keep up with adversary actions. This is especially true if each side fears that the other will act and react on ever shorter timelines. Such time pressures could help create a crisis or exacerbate an ongoing one. These pressures could arise even if the machines perform whatever analytical task is being asked of them “better” than a human would.47

Second, decision-support AI could generate bad information that leads policymakers to wrong decisions. AI systems can be wrong in opaque and unpredictable ways.48 Without careful attention and nuanced knowledge of how AI systems function, including their flaws, policymakers could end up basing consequential strategic decisions on fabricated, distorted, or otherwise shoddy information and analysis. Moreover, some scholars believe that the closed information systems of authoritarian regimes such as China make them more susceptible to bad decisions based on information from AI systems.49 Even if political-level deliberations on foreign and security policy remain relatively free of AI augmentation, similar information distortion could take place at the tactical level of military operations, leading to accidental escalation even if humans retain ultimate discretion over lethal force in areas where the U.S. and Chinese militaries interact.50

Finally, military AI could affect the decision-making and information domain by revolutionizing large-scale influence operations that inject unpredictable elements into political crises. Convincing, realistic AI-generated

Pathway 2: AI’s Effects on the Decision-Making and Information Domain

Military AI tools could increase strategic risks emanating from the decision-making and information domain in three main ways: by compressing the time policymakers have to make high-stakes decisions, by generating bad inputs to decision-making processes, and by tempting actors to try to undermine states’ deliberations through large-scale information operations. First, AI-augmented decision-making could on its own speed up the pace of events and compress timelines for leaders to act in uncertain situations. One speculative application for AI systems is to supplement existing national security processes through which evolving situations are monitored, options developed, and decisions made about the threat or use of force.

For any one state on its own, accelerating information processing could help buy additional time for humans to make more accurate and informed decisions. The net effect, however, of multiple states compressing their decision

PLA soldiers march in Moscow for a military parade on June 24, 2020, marking the 75th anniversary of the Soviet victory in World War II. The PLA is racing to build and field high-end military technologies, including those powered by AI. One challenge Beijing will face, though, is finding and retaining enough qualified personnel to operate such complex systems effectively. (Pavel Golovkin/Pool/AFP via Getty Images)
text, audio, images, or video at scale could empower greater use of mis/disinformation operations. These could be used to undermine an adversary’s political unity, confuse their decision-making, create rifts with their allies, or tilt the opinion of third-party actors against them. Many Chinese strategists already see such propaganda as a vital tool to undermine enemy cohesion and initiative. And PRC party-state organs such as the Central Propaganda Department and the United Front Work Department expend great effort on internet influence campaigns designed to seed favorable narratives among key constituencies abroad. For its part, the PLA emphasizes “cognitive domain operations,” which “take people’s will, belief, thinking, and psychology as direct combat targets, and seek to affect decision-making and actions by changing the opponent’s cognition.”

Doing so could manifest as using inflammatory message campaigns to exploit political divisions over U.S.-China relations or national security writ large, or perhaps as generating fake news reports to cast doubt on an administration’s case for acting in a crisis. Research suggests that Beijing has already begun using AI to enhance its practice of information operations surrounding Taiwan’s elections. The use of large language models (LLMs) like the now-ubiquitous ChatGPT could lower the cost and may increase the persuasiveness of such activities. Indeed, in the wake of generative AI’s popularization, many PLA strategists worry that the United States’ lead in the category could enable it to conduct damaging information campaigns against China. However, the effects of a given campaign will likely be unpredictable, as the impact of even flawlessly generated text still depends upon an accurate diagnosis of the relevant societal cleavages and upon often-volatile public opinion. If AI incentivizes either side—most likely Beijing—to more frequently wade into these murky waters, it could provoke or exacerbate political crises.

Pathway 3: Uncrewed Autonomous Systems
One of the main military applications of AI could be to enable uncrewed systems—often called “drones”—to operate autonomously. The authors use “robotic systems” as shorthand for those that can carry out their missions largely without human input, while recognizing that in practice systems may exhibit a spectrum of autonomy. Sophisticated and numerous military robots create increasing potential for deliberate, inadvertent, or accidental escalation, especially in times of crisis. The United States and China are developing such systems, both in familiar configurations, such as aircraft and submersibles, as well as in more novel ones, such as “swarms” of small robots. Artificial intelligence, whether machine learning or other computational techniques such as control theory, is a key component of functional autonomy.

Robotic systems can lead to deliberate escalation in four main ways. First, if autonomy provides superior capability, political and military leaders may simply be more inclined to use force because they believe their chances of success on the battlefield are higher. Second, the lower expected risk, in human casualties, of military operations using robots could make leaders on either side more likely to undertake them. Paradoxically, the target country similarly could be more likely to respond with force. When leaders perceive a cap on the severity of violence, it may make them more risk-acceptant in operating and responding to robotic systems in ways they otherwise would not with crewed platforms. Recent circumnavigations of Taiwan by Chinese combat and reconnaissance drones provide an illustrative example of this dynamic.

Third, increased computational power could also augment the autonomous capabilities of existing systems in ways that may lead to crisis instability. For example, military AI can help improve hypersonic weapons’ ability to maneuver in the terminal phase to avoid air and missile defenses. On the other hand, machine learning could also heighten the predictive power of air and missile defenses, enabling the deployment of counter-hypersonic and other high-end missile defense systems. More broadly, analysts debate the degree to which hypersonic weapons are destabilizing and provide truly new capabilities to the countries that operate them. All things considered, it is too soon to be certain about the net impact of AI-enabled autonomy on military-technological concepts such as cost-exchange ratios, first-strike advantages, and costs of power projection across various domains. But policymakers will have to closely watch the development of military AI for signs that it might push these variables in destabilizing directions.

Fourth, low-cost uncrewed aircraft operating in swarms could theoretically create new options for conventional counterforce strikes against an adversary’s nuclear arsenal. This potential capability contributes to a larger trend that threatens to upset the strategic balance: Chinese analysts frequently express concerns about the possibility of effective U.S. nuclear counterforce strikes, including with advanced conventional weapons such as drone swarms. They see that possibility as undermining what they call the “asymmetric strategic stability” between the United States and China by threatening Beijing’s second-strike retaliatory capability.
And they contend that Beijing’s nuclear buildup is, at least in part, aimed at restoring balance in that area.

Next, related dynamics could lead to inadvertent escalation—that is, escalation arising from an action taken intentionally but that has unintended effects. Because robotic systems are still relatively new and expectations around them are evolving, countries may harbor misperceptions about whether an adversary views a given action as crossing some escalatory threshold. If one state shoots down or seizes an enemy asset under the assumption that its adversary views such an act as part of the normal push and pull of international rivalry, but the adversary instead views it as a provocation or even an act of war, the first state will have inadvertently crossed an important strategic line.

At first blush, the historical record suggests that an incident involving an uncrewed system is unlikely to lead to serious escalation. In the past several years, both Iran and Russia have shot down U.S. drones; in neither instance did the United States retaliate with force. In these examples, both states seemingly treated an uncrewed aircraft as less of an escalatory catalyst than a crewed aircraft would have been. The United States and China have confronted a similar, if less drastic, situation: In December 2016, China seized a U.S. Navy uncrewed underwater vehicle (UUV) in the South China Sea. That incident was resolved swiftly and without escalation. In addition, China has used drones as part of its larger campaign to harass Taiwan’s military forces and contest Taipei’s administration of the territory, waters, and airspace of Taiwan and its outlying islands. Taiwan military personnel have shot down such civilian drones without provoking immediate military retaliation.

However, the above examples should not lead to complacency, as they lack important features of plausible U.S.-China crisis scenarios. Neither the Iranian drone nor the U.S.-China UUV incident included armed systems. In the Russia case, while the drone concerned was capable of attack operations and the incident happened during an ongoing political crisis, it did not involve armed systems or deliberate military actions that could have led to escalation.

The People’s Liberation Army’s WZ-7 high-altitude reconnaissance drone sits on display before the 13th China International Aviation Aerospace Exhibition on September 27, 2021. Military and civilian actors in China have invested significantly in drone technology over the past decade, making it one of the biggest exporters of the technology. (Noel Celis/AFP via Getty Images)
Pathway 4: Intelligence, Surveillance, and Reconnaissance

Military AI is already enabling new tools for fulfilling intelligence, surveillance, and reconnaissance (ISR) missions and will likely do more in the future. Military AI can be combined with old technologies to do new things or do them better or more cheaply. Possible examples include combining AI with balloons or microsatellite constellations to conduct surveillance in “near space” or employing swarming capability for ISR drones. When norms around such surveillance technologies do not exist or are flimsy, crises may arise. This happened earlier in history with the advent of high-altitude aircraft and satellite ISR in the 1950s and 1960s, despite attempts by the Eisenhower administration to reach an understanding on those capabilities with the Soviet Union. It was not until the waning days of the Cold War that the United States revived the idea of international agreement on aerial reconnaissance, leading to the 1992 signing of the Treaty on Open Skies.

Next, some improvements in military AI could beget capabilities that would change structural elements of the military-technological landscape. AI systems could enable large-scale processing of data from various sensors to track mobile missile systems on land and even submarines at sea, especially if combined with other emerging technologies, such as quantum sensors. Those applications are still only theoretical but could be feasible in the medium term. If they come to pass, they would create transparency with destabilizing effects. They could undermine the survivability—the property of a military system that makes it hard for adversary forces to find and destroy—of components of two legs of the nuclear triad by enabling adversary tracking and targeting of those assets for counterforce strikes. To be sure, some analysts contest those predictions.

Still, one real possibility could be that military AI has the highly destabilizing impact of suddenly making vulnerable capabilities that were developed and fielded precisely because of their survivability. Fear of vulnerability to a sudden first strike can create use-it-or-lose-it pressures that increase the risk of nuclear escalation. Indeed, Chinese scholars have noted this trend, along with the potential challenges to second-strike capabilities posed by conventional AI-enabled drones. Chen Qi and Zhu Rongsheng of the Center for International Security and Strategy at Tsinghua University explain that China and Russia “fear that [the United States’] powerful reconnaissance capabilities could mature into a threat to their more sophisticated retaliatory forces” and note that “all AI needs to do is undermine the level of retaliatory capability” to elicit a response. Moreover, perceptions or fears that an adversary could gain a particular capability are sometimes enough to create a sense of vulnerability or even precipitate countermeasures, regardless of whether those actually exist or could exist in a relevant time frame.

At the same time, more transparency could be stabilizing in some instances. AI could improve radar tracking systems and other surveillance tools. In the nuclear deterrence and arms control domains, some analysts argue that military AI could enhance stability and mitigate risks across a variety of applications. These include providing earlier warning, generating more accurate information to dispel misperceptions or correct sensor malfunctions, improving military planning and wargaming, and empowering new arms control verification tools.

Pathway 5: Command, Control, and Communications

Improvements in military AI could also unleash a revolution in command, control, and communications (C3). China is trying to capture a leading edge in C3 through its multidomain precision warfare concept. The United States likewise seeks advantage through its Joint All-Domain Command and Control (JADC2) concept. Both want to develop improved C3 that can gather and fuse data from different “sensors” and rapidly feed that data as battlefield information to commanders and then as targets for “shooters.” And they want to simultaneously degrade, disrupt, or destroy the other side’s C3. The net outcomes of a real-world interaction between those two systems—each improving its own C3 while damaging its adversary’s—will be impossible to assess in advance. That is due to the relative immaturity of these fully integrated “system-of-systems” capabilities; a lack of transparency, especially for China; and the many variables at play.
Nevertheless, AI promises improvements in several areas related to C3. AI could make cyber and electromagnetic warfare (EW) attacks more potent.\textsuperscript{80} As big data assumes ever-greater importance as both an input to and object of AI, both sides will face incentives to “poison” the adversary’s data by modifying training or fine-tuning datasets to intentionally degrade the system’s performance. That could lead to uncertain downstream behavior, or predictable malfunctions that adversaries could exploit, of AI-enabled C3 systems.\textsuperscript{81} One study found that data poisoning in dual-use systems, such as large language models, can be done with as few as 100 poison examples in a large data set.\textsuperscript{82}

### Both sides will face incentives to “poison” the adversary’s data by modifying training or fine-tuning datasets to intentionally degrade the system’s performance.

AI could also improve China’s capacity to direct attacks on the satellites that form the backbone of C3 networks. To be sure, most existing antisatellite (ASAT) capabilities do not rely on artificial intelligence. Instead, they employ conventional missiles, EW, and cyberattack techniques.\textsuperscript{83} But press reports have shown PRC researchers exploring concepts for small “hunter” satellites that could use AI to guide propulsion and steering as they move in irregular patterns to seek and destroy or disable the satellites they target.\textsuperscript{84} As both countries’ C3 and counter-C3 capabilities evolve, national leaders could face increased pressures to “use or lose” weapons during a crisis or conflict, which could lead to rapid escalation.

Another specific concern is that military AI could affect C3 systems for nuclear weapons. Nuclear early warning systems will increasingly rely on AI for fusing and rapidly analyzing data from a variety of sensors. Such systems could misinterpret the data and produce either false positives that report a missile is incoming when it is not, or false negatives that show no threats when in reality a missile is on its way.\textsuperscript{85} China’s decision to move at least some of its nuclear forces to a launch-on-warning posture could shorten the time available to double-check or reassess information about incoming threats.\textsuperscript{86}

Finally, concerns over time pressure or threats to senior levels of government with command-and-control authority might lead to nuclear capabilities, posture, or policy changes. Both countries will face pressures and incentives to devolve control over nuclear launch decisions down the chain of command to ensure those decisions can be made quickly in a crisis or even after a nuclear exchange. Taken to its logical extreme, China—or any other nuclear state—could decide to build an automated retaliatory capability. That concept is often referred to as a “dead hand” after a Soviet system of that type from the Cold War.\textsuperscript{87} Neither Beijing nor Washington appears to be developing such a capability, and their official policies rule it out.\textsuperscript{88} But the structural pressures will remain and could become more intense over time as military AI matures.

### Options for Managing Strategic Risks from Military AI

The United States will need to undertake a portfolio of measures to address the various dangers that military AI poses in the bilateral security relationship with China. Just as these sources of risk may overlap in practice, approaches in a risk management portfolio will aim at reducing multiple different drivers of instability. This section describes three major categories of strategic risk reduction options and, where applicable, outlines the steps already taken by Washington and Beijing within each line of effort.

### Competition through Obstructing China’s Military AI and U.S. Capability Development

One pathway by which AI may encourage deliberate escalation is if it provides one side with a large enough military capability advantage that the country believes it can launch a war and achieve its objectives at an acceptable cost. In almost all plausible contingency scenarios, the PRC would be the first mover given its revisionist ambitions. Therefore, one set of policy options for Washington involves preventing Beijing from using AI to tip the military balance of power decisively in China’s favor.\textsuperscript{89} To do so, the United States can attempt to deny China the technological advancements necessary to achieve the PLA’s “intelligentization” goals. Or Washington can develop the United States’ own AI capabilities to be sufficiently advanced and well-performing so that China never achieves a commanding lead.

U.S. leaders are already attempting to do both simultaneously. In October 2022, Washington enacted far-reaching prohibitions against exporting...
advanced computing and semiconductor manufacturing items and know-how to China. The United States has also imposed sanctions and export controls on entities and individuals that work with the PLA, including on military AI issues (in addition to other reasons). Washington places some restrictions on investments into the United States and is expected to restrict some outbound investments into China soon. Taken together, these actions are aimed at constraining China’s ability to develop military AI by stymieing commercial and industrial developments on which the PLA could draw. At the same time, the Department of Defense (DoD) is proceeding with efforts to develop and deploy military AI that can increase its overall combat power.

So far, U.S. efforts to deny China the fruits of military AI have focused on compute—the advanced semiconductors that process data for AI systems—while addressing other basic building blocks of AI, including data, human talent, and algorithms, using far more targeted tools. However, Washington may in the future place limits on these other categories. For instance, the movement underway to ban TikTok from operating in the United States is driven in part by concerns that Americans’ data could be used to fuel Chinese AI advances. Washington could also move in the future to prevent data that would be more relevant to military AI from flowing to China. The U.S. government has already placed limited restrictions on exporting or disclosing the source code of AI algorithms designed for geospatial analysis and has considered similar rules for facial recognition software in view of how China has used that technology in human rights abuses.

In the future, policymakers could move to limit the export or disclosure of general-purpose algorithms such as large language models. For example, the Committee on Foreign Investment in the United States (CFIUS) would almost certainly reject any attempt by a Chinese entity to invest in OpenAI. Washington may also fluctuate in its openness to PRC nationals’ participation in the U.S. AI research enterprise. In 2020, the Trump administration issued new rules authorizing the State Department to deny visas to Chinese graduate students with links to the PLA. The policy likely applied to roughly 2 to 3 percent of Chinese graduate students in the United States during the 2019–2020 academic year, but potentially indirectly affected a larger proportion of new enrollments in science, technology, engineering, and mathematics (STEM) graduate programs. While the Biden administration has made efforts to enable more recruitment of STEM talent from abroad, it has also retained the discretion to deny PRC applicants’ visas. While this authority does not target AI specifically, a future government could take further steps to prevent Chinese students from studying AI in the United States.

However, such a move could limit U.S. AI research efforts, due to the strong representation of PRC nationals among U.S. postgraduate AI students and the extensive, ongoing level of research collaborations between the two countries. Indeed, human capital could be a powerful point of leverage for China: more top AI researchers, even among those who work in the United States, hail from China than from anywhere else. Beijing has expended great effort in the past two decades to lure top scientific talent to China, and according to the Organisation for Economic Co-operation and Development (OECD), in 2021 it recorded a net inflow of more than 2,000 published scientists, while the United States saw a net outflow. In particular, 2020 saw an acceleration of Chinese scientists leaving American institutions for ones in the PRC, reportedly due to a combination of U.S. mismanagement of COVID-19, attendant anti-Asian racism, and harassment of Chinese-origin scientists by the U.S. government under the guise of research security.

The data lags, and trends might have reversed given shifts in American politics, China’s strict zero-COVID policies through late 2022, and Beijing’s crackdown on the tech industry. Still, in the future, especially as its own AI ecosystem continues to improve, China could try to tilt human capital flows in its direction by taking steps to dissuade or prevent its top AI talent from studying or working in the United States or allied countries. In general, overly broad attempts to obstruct China’s progress on military AI threaten to hamper the United States’ own AI innovation ecosystem, so policymakers will have to strike a difficult balance.

**Unilateral Responsible Management**

Military systems malfunctioning or behaving other than as intended, whether through human or machine error, can lead to accidental crises or escalation. Official statements by both Washington and Beijing have identified civilian casualties as a particularly dangerous effect and called for minimizing them as a key design principle for military AI. As with any weapon, the best way to reduce this risk is to give systems’ safety and reliability as much weight as their lethality or efficiency and to rigorously
carry out test, evaluation, verification, and validation (TEVV) processes on a continual basis. These principles are especially important with respect to military AI systems. That is because performance on novel or out-of-sample tasks can fail unintuitively and unexpectedly, and because the behavior of the system may continue to evolve throughout its life cycle as it accumulates more and more data.

To minimize uncertainty, the United States and China will need to adopt safe design principles and then credibly communicate that they are doing so. The United States has articulated a raft of unilateral declaratory policies about its development and use of military AI. The DoD’s 2020 Ethical Principles for Artificial Intelligence calls for AI used in the U.S. military to be “responsible, equitable, traceable, reliable, and governable.” These core principles have been reiterated and fleshed out in subsequent documents—such as the Responsible AI Guidelines in Practice, the Responsible AI Strategy and Implementation Pathway, and the January 2023 directive on Autonomy in Weapons Systems—that govern how AI is to be handled and integrated across the life cycle of defense programs.

Moreover, in February 2023, the State Department issued a declaration calling for countries around the world to adhere to similar principles and to increase transparency about their military AI development. This document also reiterated the U.S. commitment laid out in the 2022 Nuclear Posture Review to maintain, in all cases, “human-in-the-loop” control over all aspects of nuclear weapons systems. In addition, a bipartisan, bicameral group of U.S. legislators has introduced a bill to codify this policy into law.

Declaratory policy or even robust diplomacy cannot always overcome deep-rooted suspicions or remove the uncertainty from military interactions in the field.

Expert discussion in working groups for this project emphasized that both Washington and Beijing are skeptical of any proclamations of self-restraint by the other side. In light of persistent uncertainty, policymakers could consider tying their hands in designing AI into (or out of) specific military systems. For instance, at times U.S. military officials have suggested that various systems, such as the Navy’s next-generation fighter aircraft and some military sealift vessels, could be “optionally crewed” depending on the mission. In the absence of a clear understanding that rules of engagement for crewed and uncrewed systems are equivalent, China could target an aircraft with crew aboard—mistakenly believing it to be uncrewed—and end up harming U.S. personnel, thereby inadvertently escalating a situation. Washington will have to consider the potential for that type of scenario as it decides whether to develop and build certain ambiguity-inducing configurations such as optionally crewed aircraft. Creating indicators to designate the current status of an optionally crewed aircraft might be possible. And, in the end, optionally crewed aircraft might confer operational benefits that outweigh the strategic risks. But U.S. officials must grapple with the risks directly rather than dismiss them outright.

**Bilateral and Multilateral Diplomacy**

Another way of preventing dangerous power imbalances, costly arms racing, or miscalculations is bilateral and multilateral diplomacy. Through negotiated arms control agreements or confidence-building measures (CBMs), states can attempt to set boundaries for the development or use of specific military technologies and then verify compliance. Broadly limiting AI development as such is difficult, perhaps even infeasible, and probably undesirable. Artificial intelligence comprises many diffuse, dual-use technologies. Pretrained algorithms, such as Meta’s LLaMA model that leaked on 4chan shortly after it was announced, are nonrivalrous and hard-to-track digital goods. Moreover, the field is rapidly evolving. Even specific applications pose difficulties to the monitoring and verification regimes necessary for successful arms control. Nevertheless, the United States and China can and should discuss limits on the most dangerous applications of AI, such as regulating its use in nuclear command and control, or offensive cyber operations.

While the two countries have no history of bilateral arms control negotiations, both Washington and Beijing participate in ongoing discussions in the Group
of Governmental Experts (GGE) on limiting Lethal Autonomous Weapons Systems (LAWS), which has met formally since 2014 in the framework of the U.N. Convention on Certain Conventional Weapons (CCW). Neither side has called for or committed to a prohibition on developing such systems, although Beijing has previously suggested it might back the idea before reversing course. However, China’s 2021 Position Paper on Regulating Military Applications of Artificial Intelligence called in vague terms for countries to “develop and apply AI technology in the military field in a prudent and responsible manner, refrain from seeking absolute military advantage, and prevent the deepening of strategic miscalculation” and asserted that “military applications of AI shall never be used as a tool to start a war or pursue hegemony.”

Even absent binding agreements, CBMs can be an important diplomatic tool for establishing baseline expectations. Washington and Beijing can use bilateral and multilateral channels to exchange both high-level and domain-specific views about the national security implications of artificial intelligence. Ideally, the two sides would engage in military-to-military dialogues in which each side could ask questions about capabilities and their employment and communicate expectations about rules of engagement, operational deconfliction, and other topics. However, given the sensitivity of the subject and the PRC’s propensity to cancel military-to-military engagements in response to other bilateral disagreements, neither side can count on sustained dialogue between the PLA and the U.S. military.

In light of this fact, the countries could draw on Track 1.5 and Track 2 dialogues in addition to official channels. Alternatively, European allies or other third parties could play an important convening role, given the reticence of Chinese officials to engage directly with American counterparts. In April of this year, NATO Secretary-General Jens Stoltenberg put forward the possibility that the Western alliance could begin just such a dialogue, although it remains to be seen if China will view NATO as the right platform for discussions. Because crises almost always involve aberrant or unexpected situations, in an ideal world Washington and Beijing would have standing crisis communication mechanisms in addition to regular peacetime discussions about AI. The latter would theoretically furnish both personal relationships and baseline understandings to facilitate the former.

**Recommendations for Policymakers**

The emergence of military AI will likely deepen U.S.-China rivalry and increase strategic risks. In response to these trends, U.S. policymakers should:

*Take bold action to constrain China’s progress in AI for military and repressive purposes, but do so in a narrow way that avoids self-defeating steps.*

U.S. policymakers should continue to aggressively restrict semiconductor production equipment and know-how, as well as end products such as cutting-edge chips, that contribute to Beijing’s advancements of military AI and its domestic repression apparatus. In addition, Washington should seek out creative tools to regulate other basic building blocks of AI—data, algorithms, and human capital—where it is clear they are supporting such malevolent purposes. In doing so, the United States should restrict itself to technologies with clear military (and dual-use) and repression applications and continually refine its policies to ensure their effectiveness while avoiding overly broad restrictions that end up being self-defeating. The reasons for action are clear: China succeeding in its military AI ambitions and gaining a sizable advantage in the commanding heights of military-technological power threatens to make an already-serious security threat worse. And PRC repression shocks the moral conscience of the world.

There are, however, also practical reasons to keep restrictions narrow: Some actions could undermine the U.S. AI ecosystem and therefore America’s ability to compete technologically, especially by blocking the flow of talent. Most restrictions will be ineffective without the cooperation of allies and partners, many of whom are skeptical of broad-brush remedies. U.S. companies could lose out on valuable commercial opportunities, only to be backfilled by foreign competitors. Washington could remove potential sources of leverage over Beijing and lock in an even more adversarial dynamic where China concludes that the United States is trying not just to cap its military progress but
also its economic prospects. Striking the right balance in practice between these two overarching aims will often be excruciating. And even this report’s authors disagree on the wisdom of certain tactics. But balance is nevertheless essential given both the imperatives for action and the risks of overreach.

**Build U.S. military AI capabilities to stay on the cutting edge.**

The United States will have to move quickly—perhaps with even more speed than it is now and matched with commensurate TEVV—to keep up with China’s progress on developing military AI. That will require difficult reforms in numerous areas. Detailing them exceeds the scope of this report, but key areas for the DoD, military services, and U.S. Congress include changes to the acquisition system and, in some places, military service cultures that prioritize legacy weapons systems. In addition, the DoD should prioritize resilience as a critical attribute for new military systems, which will bolster deterrence and reduce “use-or-lose” pressures during crises. Success in this area will require action that goes beyond the DoD to include the need to update immigration and education policies to attract, train, and retain the best scientists and engineers from around the world.

**Develop, promulgate, and implement norms and best practices on responsible military AI.**

Washington should position itself as the world’s prime mover on shaping norms and best practices for the development and operation of military AI. The United States has already taken some ambitious early steps in this regard. But Washington can go even further by generating and sharing information about TEVV processes that ensure that fielded military systems perform in ways that comport with the overarching principles outlined in those documents. Key near-term priorities for the United States should include further fleshing out the operational specifics for implementing norms against cyberattacks (including AI-enabled) on nuclear C3 infrastructure, and for fulfilling the U.S. Nuclear Posture Review’s promise to “maintain a human ‘in the loop’ for all actions critical to informing and executing decisions by the President to initiate and terminate nuclear weapon employment.” Concrete U.S. steps to implement these principles and ensure that military AI is robust, reliable, and effective will be essential for any of these norms and best practices to have credibility—especially with a skeptical China. Simply put, U.S. actions must match its rhetoric on responsible use of military AI.

**Proactively engage with like-minded allies and partners and in multilateral institutions on military AI issues.**

U.S.-China competition, including over military AI and related security issues, does not happen in a vacuum. The regional and global context plays a vital role in shaping outcomes. Washington should proactively build consultations on these issues into its alliance and partner relationships. Early discussions of military AI in NATO, the AUKUS partnership, and bilateral alliances with Japan and South Korea should be expanded, including potentially into the G7 grouping. These forums offer constructive places to hash out key tactical questions with like-minded partners, such as whether agreements should regulate specific technologies or instead focus on regulating certain outcomes. They can also provide platforms for clarifying where and how existing international laws already cover military AI issues. In addition, Washington should continue to proactively advocate for its position in multilateral forums, including with the LAWS process mentioned earlier. These activities will help shape China’s options related to military AI. Beijing is more likely to align with military AI principles and practices supported by a coalition of global actors than ones pushed solely by Washington.

**Negotiate risk reduction and confidence-building measures with China related to military AI.**

Beijing has been unwilling to engage in substantive strategic risk reduction talks related to nuclear weapons. Chinese officials have cited the disparity in their arsenal sizes to justify that stance (although China’s nuclear modernization is shrinking the gap). But Washington should test the possibilities for a channel on military AI given that the militaries are much more evenly matched in their capabilities in that area. Early activities should focus on producing relatively basic outcomes such as developing a glossary of military AI terms and their U.S. and Chinese equivalents. That would help ensure both sides have common definitions for key concepts, which would facilitate diplomacy and reduce misperceptions that could come from language and cultural barriers.

The two sides could also develop a hierarchy of risk tiers associated with different capabilities. For example, AI for logistics and maintenance is low risk, while AI-enabled autonomous nuclear weapons would be extremely high risk. The two powers could further discuss how and where both sides are, and
are not, employing AI for military purposes as well as norms and expectations around AI’s role in the use of lethal force. Even if U.S. and Chinese officials disagree, talking about these issues can help increase mutual understanding and reduce the risks of miscalculation and misperception. A more ambitious objective might be to negotiate incidents-at-sea and incidents-in-air agreements and/or agreements related to rules of engagement for uncrewed autonomous systems.

Continue to pursue universal U.S.-China risk reduction and crisis management mechanisms despite persistent challenges.

As both superpowers integrate military AI into their forces, the task of forging effective U.S.-China diplomatic channels for reducing strategic risks and managing crises that arise—what the Biden administration frequently calls “guardrails”—will become more critical. Unfortunately, U.S.-China strategic risk reduction and crisis management mechanisms have a poor track record overall, although arguably there have been a few successes. Still, even intermittently functional risk reduction and crisis management mechanisms provide more value than none at all—provided Washington does not make meaningful policy concessions just to keep meeting. Leader-level contacts in particular offer a chance to communicate directly, which can reduce the misperceptions and misunderstandings that come from having messages filtered through intermediaries and staff. And especially in the Chinese system, Xi’s support for continual-if-difficult diplomacy is required for lower-level meetings to proceed.

Make military AI a fundamental pillar of diplomacy with China related to nuclear weapons and strategic stability.

Nuclear risks are rising in the U.S.-China relationship. Multiple factors are driving that trend, including the rapid expansion in both size and sophistication of Beijing’s nuclear arsenal. As this report has shown, military AI also plays an increasingly important and potentially destabilizing role in shaping the balance of nuclear and other strategic capabilities. Biden and Xi reportedly agreed in their November 2021 meeting that their two sides would “begin to carry forward discussions on strategic stability.” Unfortunately, there is no public evidence those talks ever got off the ground. Washington should instead propose to restart that process in the permanent five (P5) group of recognized nuclear weapons states and then build military AI into the negotiations. Chinese officials have themselves previously endorsed this idea and are therefore more likely to engage in a meaningful way. And earlier the group produced an important P5 statement on nuclear issues (although Russian nuclear threats over Ukraine have undermined the credibility of Moscow’s signature).

Take steps to reduce strategic risks not directly caused by military AI but potentially worsened by the inherent speed and unpredictability of military AI.

U.S.-China relations are likely to grow more volatile and uncertain as military AI matures. Recognizing that fact, Washington should take steps to reduce strategic risks in other areas where possible. Unilateral operational decisions that show restraint, such as delaying ICBM tests during moments of heightened tensions, as the United States did in August 2022, are wise, particularly when tests are not immediately necessary to ensure a safe, secure, and effective nuclear deterrent. The U.S. initiative to create norms that ban destructive, direct-ascent antisatellite missile testing provides another positive example. In the future, the United States and China could ink a missile launch notification agreement similar to the pacts that each has with Russia.

Prioritize intelligence-gathering and analysis on, and net assessment of, China’s military AI capabilities.

This report has shown that it is possible to offer some preliminary analysis of how military AI will affect U.S.-China security relations and potentially worsen strategic risks. But the full trajectory and effects of military AI cannot be known yet. Military AI could fizzle and end up being less important than current projections expect—or it could revolutionize military affairs in unimaginable ways. Moreover, changes will be determined not just by military AI itself, but how it interacts with other types of factors. These include nuclear arsenals and associated infrastructure, conventional capabilities, diplomatic relations between the United States and China, and developments in civilian AI technologies.

Therefore, there is a pressing need to deepen understanding of how military AI might, or might not, contribute to arms racing dynamics, the potential for crisis escalation, and strategic stability overall. The U.S. director of national intelligence, secretary
of state, and secretary of defense should task their organizations with refining or, if needed, building multidisciplinary offices and cadres of experts to monitor, analyze, and proactively recommend policy changes related to this problem set. Each of those offices should have formal linkage to, or overlap with, the components within each department—often called “China Houses”—tasked with coordinating policy related to China across regional and functional issues.

## Conclusion

The advancement of military AI threatens to inject additional tensions and instability into the U.S.-China security relationship. Assessing and managing the resulting strategic risks will require keeping a close watch on Beijing’s civilian and especially military AI activities. Improving understanding of how military AI might create new pathways for crisis or conflict as new capabilities come online will likewise be essential for analysts and policymakers. In this context, reducing strategic risks will ultimately require a combination of competition and selective engagement with China. As U.S.-China relations struggle to find a stable equilibrium and AI capabilities leap ahead, the need for policymakers to grapple with the challenges presented by the interaction between these trends will continue to grow.


3. This definition of “strategic risks” draws from language in “U.S.-Russia Presidential Joint Statement on Strategic Stability,” The White House, press release, June 16, 2021, https://www.whitehouse.gov/briefing-room/statements-releases/2021/06/16/u-s-russia-presidential-joint-statement-on-strategic-stability/. Many analyses on similar topics use the term “strategic stability,” but that concept has many competing definitions, including ones that apply the concept narrowly to certain aspects of the nuclear correlation of forces. The authors therefore concluded that it is clearer to use the general term “strategic risks” to describe the areas covered in this report.

4. Because this report is written for U.S. and allied policymakers and scholars, and given the authors’ areas of expertise, we do not provide an in-depth analysis on the United States’ approach to these issues. For a good primer, see Paul Scharre, Army of None: Autonomous Weapons and the Future of War (New York: WW Norton & Company, 2018); and Ben Buchanan and Andrew Imbrie, The New Fire: War, Peace, and Democracy in the Age of AI (Cambridge: The MIT Press, 2022).


19. The European Union’s proposed AI Act, for instance, seeks to both regulate the largest and most powerful AI models and draw strict limits around specific AI applications such as facial recognition and predictive policing. Natasha Lomas, “EU lawmakers back transparency and safety rules for generative AI,” TechCrunch, May 11, 2023, https://techcrunch.com/2023/05/11/eu-ai-act-mep-committee-votes/.


42. For example, the PLA recently tested an AI system to help improve targeting for long-range artillery. Stephen Chen, “China tests AI-powered long-range artillery that can hit a person 16km away,” South China Morning Post, April 17, 2023, https://www.scmp.com/news/china/science/article/3217334/china-tests-ai-powered-long-range-artillery-can-hit-person-16km-away.

43. Ding and Dafoe, “Engines of Power.”


47. It is beyond the scope of this paper to describe in a general way what it means to evaluate strategic decision-making as “better” or “worse.” The authors refer narrowly here to AI systems making decisions that are more effective as measured by the parameters for which the system is designed to optimize.


56. Some studies have shown that influence operations (even, in one study, simulated pro-Beijing messages about Taiwan) can have significant backfire effects. Jon Bateman et al., “Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research” (Carnegie Endowment for International Peace, 2021), https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824.

57. To be sure, not all uncrowed systems are autonomous or even semi-autonomous, and the authors recognize the important distinction between flying, sailing, and operating autonomously, and having the autonomy to use lethal force.

58. There is a robust debate about the practical and ethical considerations surrounding the use of lethal force by autonomous machines—the so-called killer robot question. Many of them center on whether a human has final authority over lethal force, and how that authority is implemented. This debate is too multitudinous to effectively summarize here. Rather, we aim to identify pathways by which autonomy can lead to crisis instability regardless of whether they are designed to independently use lethal force against humans.


97. Scharre, Four Battlegrounds, 30–34.


104. U.S. Department of State, “Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy.”


109. At present, training leading-edge machine learning models requires enough concentration of specific human and material resources that it is hard to hide. But as hardware and software progress and know-how proliferates, the cost of training models, and thus the visibility of the process, will likely decrease.


119. Both authors agree that it is geopolitically and normatively wise to try to keep the building blocks of powerful AI away from the Chinese military and entities with a documented history of enabling systematic human rights violations. They part ways on the breadth with which one should apply this standard in the context of China’s political economy. In general, Stokes believes that the combined risk of China achieving a military AI breakthrough and commercial technology making its way to the PLA justifies measures such as the broad chip export bans of October 2022. Sullivan judges that such measures will fail to achieve their objectives while reducing the United States’ leverage and tarnishing its reputation, among other pitfalls.


121. For an early exploration of this topic, see Flournoy, Haines, and Chefitz, “Building Trust through Testing.”


124. To be sure, PLA and China Coast Guard forces have a track record of ignoring existing agreements for operational safety when doing so gives them a tactical advantage. Despite that, U.S. policy is to continue to operate American forces according to these professional guidelines to demonstrate responsibility and adherence to a rules-based security order. The same pattern would likely recur or persist even if Washington and Beijing concluded new agreements in the field of military AI.


128. In January 2022, Fu Cong, director-general of the Chinese Foreign Ministry’s Arms Control Department, told reporters that the P5 countries should talk “more directly” about global security. He went on to add that, “Strategic stability goes beyond nuclear. … Our idea is to expand the subject of the P5 process so we could discuss not only the nuclear issues, but also other issues related to strategic stability, including outer space, missile defense, even AI and other emerging technologies.” As quoted in Bloomberg News, “China Calls on Nuclear-Armed Nations to Focus on AI, Space,” January 4, 2022, https://www.bloomberg.com/news/articles/2022-01-04/china-calls-on-nuclear-armed-nations-to-focus-on-ai-space.


About the Center for a New American Security

The mission of the Center for a New American Security (CNAS) is to develop strong, pragmatic and principled national security and defense policies. Building on the expertise and experience of its staff and advisors, CNAS engages policymakers, experts and the public with innovative, fact-based research, ideas and analysis to shape and elevate the national security debate. A key part of our mission is to inform and prepare the national security leaders of today and tomorrow.

CNAS is located in Washington, D.C., and was established in February 2007 by co-founders Kurt M. Campbell and Michèle A. Flournoy. CNAS is a 501(c)3 tax-exempt nonprofit organization. Its research is independent and nonpartisan.

©2023 Center for a New American Security
All rights reserved.