Research on Artificial Intelligence and International Security 2026

Development, Application, and Governance of Artificial Intelligence in China and the U.S.

June 2026

清华大学战略与安全研究中心
CENTER FOR
INTERNATIONAL SECURITY AND STRATEGY
TSINGHUA UNIVERSITY

Introduction

In May 2026, during their meeting in Beijing, the heads of state of China and the United States held constructive exchanges on artificial intelligence and agreed to launch an intergovernmental dialogue on AI. As two major AI powers, China and the United States should work together to advance the development and governance of artificial intelligence, so that AI can better serve the progress of human civilization and the shared well-being of the international community. Since October 2019, the Center for International Security and Strategy (CISS) of Tsinghua University and the Brookings Institution have jointly held the China-U.S. Track 2 Dialogue on Artificial Intelligence and International Security, conducting a total of 14 rounds of discussions on issues related to artificial intelligence. Each side released phase-based project research reports and three research reports on AI-related terminology, continuously deepening mutual understanding over key issues. Against the backdrop of China and the United States announcing the resumption of the intergovernmental dialogue on artificial intelligence, researchers from the CISS AI project team have drawn on insights from previous dialogue sessions. They examine the development, application and governance of artificial intelligence in China and the United States across themes including AI safety and cooperation, the development of artificial general intelligence, military applications, and disinformation governance, offering reflections for sound interaction and joint progress between the two countries in the field of artificial intelligence.

Contents

AI Safety Is Where U.S.-China Cooperation Still Matters

XIAO Qian

As discussions grow around the upcoming visit by U.S. President Donald Trump, much attention has focused on tariffs, trade, and semiconductors. Many expect that artificial intelligence will also feature prominently on the agenda. But if AI is discussed only through the lens of technological rivalry, both countries may miss one of the few remaining areas where meaningful cooperation is still possible - and urgently needed.

In recent years, Washington's approach to AI has increasingly shifted from innovation governance towards strategic competition. The language of "winning the AI race" now dominates many policy discussions in the United States. Export controls, investment restrictions and technology alliances are increasingly framed through national security considerations.

It is true that AI is rapidly becoming a foundational technology with major economic, military and geopolitical implications. No major power wants to fall behind. Yet the problem is that AI is not only a competitive technology. It is also a technology that may cause systemic risks.

Some of the most serious risks associated with advanced AI do not stop at national borders and cannot be managed by one country alone. Whether in AI-enabled cyber operations, biological research, autonomous military systems or large-scale disinformation, instability created in one country can quickly spill across the international system. This is why AI safety should become an important topic during any future high-level engagement between China and the United States.

China has repeatedly demonstrated that it is open to international dialogue on AI governance and safety. In recent years, Beijing has issued multiple policy documents and governance initiatives emphasizing the importance of balancing innovation and safety, while Chinese institutions, scientists and policy experts have actively participated in international AI safety discussions.

This point is often overlooked in the current geopolitical atmosphere. Much of the international discussion today focuses on strategic competition in AI capabilities, chips and infrastructure. But behind the headlines, there is also another reality: Chinese and American experts are still talking to each other seriously about AI risks. Despite growing tensions between the two countries, Track II dialogues involving scientists, researchers and policy experts from both sides have continued quietly over the past few years. These discussions cover issues ranging from loss-of-control risks to AI-enabled cyber threats and AI x bio risks, from terminology clarification to confidence-building measures.

I have personally observed that many experts on both sides, regardless of political differences, share similar concerns about the future trajectory of AI development. There is growing recognition that some risks are simply too large for any country to handle independently.

Both China and the United States have a responsibility to mitigate AI-related risks. As the world's two leading AI powers, decisions made in Beijing and Washington will shape not only the future of technological development, but also the global risk environment surrounding advanced AI systems.

The latest International AI safety report led by Turing Award winner Yoshua Bengio and authored by over 100 AI experts, has listed two disconcerting areas which deserve particular attention.

The first is the intersection between AI and cyber attacks. Advanced AI systems are rapidly changing the cyber domain by lowering barriers to sophisticated attacks, accelerating automated vulnerability discovery and increasing the scale of information manipulation. In an environment already characterised by deep mistrust between major powers, AI-enabled cyber incidents could easily be misinterpreted as intentional escalation.

The second area is AI and biological risks. AI is already accelerating scientific discovery in medicine and biology in remarkable ways. But many scientists also worry that increasingly capable AI systems could lower the technical barriers for harmful biological misuse. This concern is no longer limited to governments. It increasingly involves the possibility that small groups or even individuals could gain access to capabilities once restricted to advanced state laboratories.

Neither China nor the United States benefits from such a future. This is precisely why maintaining channels of communication on AI safety matters, even during periods of broader strategic competition. One of the underappreciated dangers of the AI era is that system failures may increasingly be seen as hostile intent. As AI systems become more autonomous and opaque, it may become harder for governments to distinguish between technical malfunction, unintended escalation and deliberate attack. In a crisis situation, this ambiguity itself could become destabilising.

History has repeatedly shown that excessive securitisation of technology can deepen mistrust and fragment the global innovation ecosystem. Today, there is a growing risk that AI governance itself becomes absorbed entirely into geopolitical confrontation. If every AI advance is interpreted primarily through a zero-sum framework, then safety cooperation becomes politically difficult precisely when it is most necessary.

A more pragmatic approach would recognise that AI safety cooperation serves the interests of both countries. China and the U.S. could expand expert dialogues on frontier AI risks, strengthen communication channels regarding major AI incidents, support joint discussions on AI evaluation and safety testing, and explore

confidence-building measures in areas related to military AI and cyber stability.

Some foundations already exist. Over the past few years, Chinese and American experts have worked together through a number of Track II dialogues on AI and international security. Scientists and researchers from both countries continue to exchange views on risk governance, and emerging technological threats. These exchanges may not attract headlines, but they remain valuable because they help reduce misunderstanding in an increasingly uncertain technological environment. At a time when many traditional channels of bilateral trust have weakened, AI safety remains one of the few areas where constructive engagement is still possible.

For Washington, however, this also requires a shift in mindset. Treating China exclusively as a strategic rival in AI may generate domestic political consensus, but it risks narrowing the space for practical cooperation on shared risks. Not every aspect of AI governance should be viewed through the framework of technological containment or ideological competition.

The real question facing both countries is not simply who leads in AI capability development. It is whether the world's two largest powers can act responsibly enough to prevent AI-related risks from escalating into crises that affect everyone. That is why AI safety is not a peripheral issue in U.S.-China relations. It may become one of the most important tests of whether cooperation between the two countries is still possible in this age of uncertainty.

(XIAO Qian, Deputy Director, Center for International Security and Strategy of Tsinghua University)

How China and the U.S. Can Expand Artificial Intelligence Cooperation

SUN Chenghao

Looking back over the past period, even as technological competition between China and the U.S. has intensified, the two sides have also made some constructive progress in cooperation on artificial intelligence (AI).In May 2024, China and the U.S. held the first meeting of the inter-governmental dialogue on AI in Geneva, Switzerland. The discussions focused on risks associated with AI technologies, global governance mechanisms and issues of mutual concern. This meeting marked the formal inclusion of artificial intelligence as a standing item in China-U.S. governmental dialogue, signaling a cautious but meaningful step toward managing emerging technological risks through direct communication.Subsequently, in November 2024, Chinese President Xi Jinping and then U.S. President Joe Biden met and reached an important consensus on the principle of ensuring that nuclear weapons are always under human control. This shared understanding drew a clear red line with respect to the potential militarization of artificial intelligence, particularly in nuclear-related domains. It also laid an important foundation for bilateral, and potentially global efforts to strengthen risk management and strategic communication in the face of rapidly advancing technologies.

Following Donald Trump's return to office, competition between China and the U.S. in the field of artificial intelligence has remained intense. Yet after the 2025 leaders' meeting in Busan, both sides stated that they would continue to advance mutually beneficial cooperation in AI. This signal suggests that, regardless of whether a Democratic or Republican administration is in power, there is a shared recognition in Washington of the necessity to maintain engagement and coordination with China on artificial intelligence-related issues.

That said, under the current political climate in China-U.S. relations, pursuing deeper and more concrete cooperation in areas related to artificial intelligence and the AI-nuclear nexus continues to face significant practical constraints. Nuclear weapons and other key strategic capabilities remain highly sensitive, making mutual transparency difficult to achieve in the near term. Challenges related to verification and persistent deficits in strategic trust in the arms control domain further complicate dialogue between the two sides. Even so, China and the U.S. still have a strong incentive to explore cooperation to the greatest extent possible in less sensitive areas.

On military-related issues, both sides could focus on promoting and deepening existing principled consensus, rather than rushing into substantive military dialogues or technical cooperation. In non-military domains, however, there remains considerable room and necessity for China and the U.S. to engage in

dialogue and cooperation on assessing technological risks, addressing ethical concerns, and guiding the development and application of artificial intelligence toward beneficial and responsible ends.The consensus that nuclear weapons must remain under human control provides a solid foundation for the next stage of cooperation, especially in managing AI-related risks in the military domain. President Trump is also trying to shape his image as a peace president, which opens space for both expansion and deepening of this consensus.

In terms of expanding cooperation, China and the U.S. can move in two directions. First, they can encourage other nuclear powers, such as the U.K., France, and Russia, to adopt the same principle, gradually transforming a bilateral understanding into a multilateral one. They could also take this issue to the UN framework and jointly promote statements reaffirming the principle of human control over nuclear weapons and emphasizing responsible deployment of military AI. Second, they could extend the principle beyond nuclear weapons to other strategic systems with high deterrence potential. The definition and scope, however, require further discussion and could begin in Track-2 dialogues.

On deepening cooperation, both sides can work on more detailed risk assessments and management mechanisms, breaking down nuclear weapon systems and conflict scenarios into deployment and decision-making steps to identify specific risks. A complete ban on AI in NC3 (Nuclear Command, Control, and Communications) systems is unrealistic, and both sides should also consider the potential benefits of AI-nuclear integration. Therefore, identifying mutually acceptable red lines is crucial.

Against this backdrop, Track Two dialogue serves as an important supplementary channel. Think tanks, scholars and retired military officers can continue discussions on risk assessment, ethical norms, and crisis decision-making. These efforts help sustain communication and build expert consensus that can eventually support official talks.

In non-military domains, China and the U.S. could also work together to advance global governance of artificial intelligence. To begin with, both sides could focus on the cross-border risks generated by AI and seek to promote risk-tiering, classification, and assessment frameworks that are acceptable to a broader range of countries, while jointly exploring possible responses.These governance challenges are shared by both countries. At the technical level, they include risks related to loss of control over advanced AI systems, the limited interpretability of large models, and so-called “hallucinations” in model outputs. At the application level, they encompass AI-related biosecurity risks, the potential for AI to “enable” terrorist organizations and other non-state actors, and the erosion of social trust caused by deepfakes. All of these issues transcend national boundaries and cannot be effectively addressed by any single country acting alone.Approaching cooperation from a risk-based perspective may also help reframe China and the U.S. from mutual “technological competitors” into “co-risk bearers.” Even under conditions of strategic competition, this reframing could enable the two sides to

identify shared concerns and practical areas for cooperation in AI risk assessment and management.

Second, China and the U.S. could engage in dialogue on the profound impact of artificial intelligence on economic and social structures. AI is reshaping labor-capital relations and transforming traditional modes of production. Across a growing number of sectors, trends toward unmanned operations and high levels of automation are becoming increasingly visible. While these developments may significantly enhance productivity, they also carry the risk of structural unemployment.A shared challenge for both countries is how to ensure that AI improves people's livelihoods rather than exacerbating disparities in the distribution of social resources. Closely related are questions of how to enhance societal adaptability to technological change through policy measures such as education reform, skills transformation, and more inclusive access to digital infrastructure. These issues represent areas of common concern for China and the United States. By taking the lead in advancing such dialogue, China and the U.S. could also offer useful reference points for other countries grappling with similar challenges, thereby contributing to broader international discussions on the societal governance of artificial intelligence.

Third, China and the U.S. could further explore how AI might be leveraged to provide new solutions for global public governance. In the field of climate governance, AI can be applied to disaster forecasting, extreme weather modeling and emergency response. In the realm of international peace and security, AI technologies may help support conflict mediation, optimize peacekeeping deployments and enhance early warning of crises. In global development, efforts to promote open-source technologies and the sharing of foundational algorithms could also contribute to narrowing the technological gap between the Global North and the Global South.Building on existing United Nations frameworks, China and the U.S. could embed cooperation on AI as a global public good into multilateral agendas. This could include launching pilot projects in areas such as climate monitoring, disaster management, public health early-warning systems, smart agriculture, cross-border infectious disease forecasting and the digitalization of basic education. Cooperation in these fields would not only highlight the positive potential of artificial intelligence, but also help inject elements of collaboration into a bilateral relationship otherwise characterized by intense technological competition, thereby contributing to a modest improvement in overall China-U.S. relations.

Finally, China and the U.S. should also jointly address the ethical and legal challenges posed by AI. The rapid development of AI technologies has brought fundamental ethical questions into sharper focus, including how to define human identity and how to preserve human agency and a sense of value in an era of increasingly autonomous systems. How to maintain effective human control over critical decision-making processes, how to establish baseline ethical principles for the development and deployment of AI, and how to clarify responsibility

and accountability mechanisms in cases of harm have all become issues of global concern. Dialogue on these questions could help further build mutual trust between China and the U.S., while also contributing to the exploration and consolidation of normative consensus within the international community on technological development.

It should be emphasized that cooperation between China and the U.S. on global AI governance is not intended to construct any form of a "G2" model. Rather, it reflects the responsibilities that arise from the two countries' respective capabilities in technological innovation, industrial capacity and international influence. Any effective framework for global technology governance cannot function without the participation of both China and the United States.As a paradigmatic disruptive technology, AI will not only profoundly reshape economic, social, and security landscapes, but will also help define the rules and governance structures of the future international order. If China and the U.S. can take early cooperative steps in areas such as risk perception, ethical principles, and public-interest applications, such efforts would carry important demonstrative value and reflect the responsibility of major powers to lead by example.

(SUN Chenghao, Fellow, Center for International Security and Strategy of Tsinghua University; Munich Young Leader 2025)

A Comparative Study of the Divergent Paths to Artificial General Intelligence between China and the United States

LU Chuanying

In recent years, Artificial General Intelligence (AGI) has re-emerged as a central theme in artificial intelligence research and technology policy debates. In contrast to “narrow artificial intelligence” tailored for specific tasks or application scenarios, AGI is conventionally conceptualized as a general-purpose intelligent system capable of cross-domain learning, reasoning and adaptation. Its potential impacts are regarded as extending far beyond individual industries or technological spheres, and will profoundly reshape economic structures, social governance, and even international power configuration. For these reasons, AGI is no longer merely an engineering or academic problem, but has gradually evolved into a political-economic issue of paramount strategic importance.

Contemporary discussion on AGI is primarily shaped by two dominant narratives. The first is the narrative of “technical feasibility” led by the technical community, which emphasizes a fundamental re-understanding of the nature of intelligence through foundational layers such as cognitive modeling and learning mechanisms. The second is the narrative of “engineering feasibility” advanced by industry, which advocates incrementally approaching AGI by continuously scaling model size, computational power and data volume, based on the existing deep learning paradigm. These two narratives each command considerable support in both academia and industry. The divergence between Yann LeCun and Demis Hassabis over whether large models can lead to AGI, for example, vividly illustrates this debate.

Yet these two seemingly opposing narratives in fact share an implicit premise: the development of AGI is primarily a process driven by the internal logic of technology and industrial incentives, with the role of the state confined largely to “support” or “regulation”. While this assumption may still hold in the era of narrow artificial intelligence, it appears increasingly inadequate in the field of AGI, which is defined by extreme uncertainty, massive investment and far-reaching externalities. AGI is confronted not with a single technical bottleneck, but with the compounding of multiple uncertainties: ambiguous technical pathways, unpredictable timing of its realization, and unquantifiable societal impacts. Under such conditions, exclusive reliance on market mechanisms not only risks inducing technological path locking and distorted resource allocation, but also proves inadequate to address the resulting societal risks and political pressures.

Against this backdrop, the development of AGI must be reconceptualized as a strategic issue. AGI possesses the distinct attributes of both a public good and

a strategic asset. Its research and development require long-term, high-risk, cross-sectoral coordination of investment, entailing the systemic integration of computational power, data, talent, capital and governance frameworks. The critical role of the state lies not in substituting the market to make technological choices, but in establishing institutional arrangements that, under conditions of fundamental uncertainty, render technological exploration economically sustainable, politically acceptable, and socially risk-controllable.

From this perspective, the divergent pathways toward AGI exhibited by China and the United States acquire considerable analytical significance. As the two global leaders in contemporary artificial intelligence development, China and the United States not only stand at the forefront in technological capacity and investment scale, but also display substantial differences in institutional structures, state-market relations, and visions of technology governance. These differences are not merely rhetorical at the level of policy formulation; they have gradually become internalized into distinct modes of advancing AGI, forming two representative institutionalized development trajectories.

The United States began systematically laying the groundwork for AGI as early as 2016, with the release of Preparing for the Future of Artificial Intelligence during the Obama administration. The Executive Order on Maintaining American Leadership in Artificial Intelligence and The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update issued during the first term of the Trump administration further reinforced the strategy of technological leadership in the AGI domain. In January 2025, the Executive Order on Advancing American Leadership in Artificial Intelligence Infrastructure leveraged the Stargate Initiative as its cornerstone, focusing on building ultra-large-scale data centers and computational power hubs to provide a continuously scalable computing foundation for next-generation general models. A comprehensive policy package consisting of four executive orders was released in July 2025, ranging from Winning the Race: An Artificial Intelligence Action Plan to measures aimed at preventing “woke AI,” accelerating data center construction, and promoting technology exports. This package sought to ensure breakthroughs in cutting-edge models while maintaining technological generational advantages through infrastructure investment and export controls. In November, the executive order Launching the Genesis Mission directly integrated frontier models into national-level scientific research projects in energy, materials, climate science, and life sciences. Its objective was to accelerate the translation of general capabilities into scientific discoveries and engineering breakthroughs through interdisciplinary computing power integration and collaborative model applications. This strategic layout relied heavily on the technological capabilities of a small number of leading enterprises, with the expectation of establishing global standards through technological leadership.

The current approach adopted by the United States can be summarized as a model-centric path. Under this model, AGI is regarded as the ultimate goal of technological innovation, and the core of institutional coordination lies in concentrating

resources to drive continuous breakthroughs in model capabilities, while shaping de facto global standards through technological leadership. In this context, the government mainly functions through strategic planning, basic research funding and safety regulation, whereas concrete technological exploration and engineering implementation rely heavily on a handful of large technology enterprises. This model carries strong potential for breakthroughs and international competitive advantages in the short run, yet its inherent risks are equally pronounced. Most notably, its technological trajectory is highly concentrated and susceptible to path locking; its extreme concentration of computing power and data resources exacerbates ethical and security concerns; and its room for adjustment becomes relatively constrained once bottlenecks emerge in key technological pathways.

China's strategic documents do not explicitly employ the concept of AGI, and its approach shares both similarities and differences with that of the United States. The two sides alike attach great importance to the technological and industrial development of artificial intelligence. Where they diverge is that China's pathway toward AGI is led mainly by application-driven innovation, rather than betting exclusively on advancing cutting-edge model capabilities to reach AGI. The 2017 New Generation Artificial Intelligence Development Plan laid the foundation for a basic three-in-one framework of technology, industry and governance, around which all subsequent policy evolution has centered. A series of application-oriented policies issued between 2022 and 2025 exhibit a distinct infrastructure-focused orientation. By opening up application scenarios, building computing power networks and enhancing cross-departmental coordination, these policies drive the systematic penetration of intelligent capabilities into the real economy and the field of social governance. The Global AI Governance Action Plan released in July 2025 and the Implementation Opinions on the Special Action for "Artificial Intelligence (AI) + Manufacturing" issued in January 2026 indicate that China is embedding the cultivation of AGI-capable competencies into its agenda of industrial upgrading and global governance.

In this process, China has gradually exhibited a "general-purpose infrastructure path". Under this model, AGI is not conceived as a singular technological end-point, but as a foundational capability that must be embedded into the national development system. Instead of wagering on the extreme breakthrough of a single technical route, the approach centers on the systematic integration of computing power networks, data systems, application scenarios and institutional mechanisms to drive the diffusion and accumulation of intelligent capabilities across a wider range of economic and social systems. The state assumes a more proactive role as overall coordinator, providing space for diverse technological pathways to thrive through infrastructure development, scenario openness and cross-departmental coordination.

These two pathways are not simply a dichotomy between state-led and market-led models, but rather represent distinct responses to the same question: under conditions of high technological uncertainty and frequent market failure, how can

institutional arrangements sustain the feasibility of long-term exploration? The U.S. model-centric path places greater emphasis on accelerating breakthroughs through market competition and technological concentration. Its strengths lie in innovation efficiency and global influence, while its risks center on instability at the technological and governance levels. China's general-purpose infrastructure path, by contrast, prioritizes mitigating overall risks through systemic integration and capability diffusion. It offers advantages in terms of stability and controllability, while simultaneously confronting challenges such as escalating coordination costs and unpredictable breakthrough timelines.

Importantly, there exists no simplistic hierarchy of superiority or inferiority between the two developmental pathways. Rather, they embody rational strategic choices made by China and the United States regarding highly uncertain technology of AGI, constrained by their respective institutional contexts, developmental stages and strategic objectives. The developmental trajectory of AGI is by no means singular; its modalities of social embeddedness, structures of risk distribution, and reverse shaping effects on national capacity will all vary drastically in line with divergent modes of institutional coordination. From a broader analytical perspective, the divergence between China and the United States in the AGI sphere extends far beyond technological competition per se, and instead foreshadows divergent paradigms for integrating general-purpose technologies into national governance systems in the future. While substantial uncertainty still shrouds the ultimate realization of AGI, one conclusion is unequivocal: the exploration of AGI has already emerged as a critical testing ground for assessing states' institutional coordination capabilities. Understanding the distinct pathways to AGI pursued by China and the United States not only facilitates a grasp of the tangible trajectory of artificial intelligence advancement, but also provides a pivotal analytical lens for deliberating on the interplay between cutting-edge technologies and national institutions.

(LU Chuanying, Professor, Vice Dean, School of Political Science and International Relations, Tongji University; Vice President, Shanghai Association for Artificial Intelligence and Social Development)

China-US Coopetition in AI's Military Applications

QI Haotian

A New Arena

Artificial intelligence (AI) is rapidly becoming an indispensable part of the military relationship between China and the United States, with both countries viewing it as a key factor in shaping the future balance of military power. The United States emphasizes accelerating the application and deployment of AI in defense and military affairs in order to preserve its advantages on future battlefields. China likewise attaches great importance to the role of AI in national defense and security. President Xi Jinping has repeatedly stressed the need to accelerate "intelligentization" in military development, and has made clear that the People's Liberation Army is to be basically modernized by 2035 and transformed into a world-class military by the middle of this century. Frontier technologies such as AI occupy an important place in that strategic objective. AI has been placed at the core of both countries' military modernization strategies and transformation agenda, and has in effect, become a new arena of China-U.S. military competition.

Yet AI does not play a singular role in China-U.S. military relations as merely a "catalyst for conflict". At a time when both countries face pressing demands for defense modernization, the impact of AI on their military relationship is multifaceted. Military applications of AI are inherently complex. They can drive revolutionary improvements in military effectiveness and changes in operational domains and areas, but they are also fraught with challenges, shortcomings, and uncertainties in delivering the empowerment they promise, while generating risks and threats that could undermine battlefield performance, national security and strategic stability. The application and deployment of AI in the military sphere is both a powerful tool for enhancing combat effectiveness and a new domain of competition between the two sides; at the same time, it can compel both countries to confront the common security risks they face. It is therefore necessary to examine China-U.S. military AI relations through the dual lenses of competition and cooperation. In addition, as two major powers of global consequence in the development of both AI and military might, China and the United States should jointly shoulder responsibility for addressing the structural security challenges created by the militarization of AI, seek dialogue and coordination mechanisms in relevant governance areas, and prevent technological and operational loss of control from harming national security, disrupting regional and global stability.

Empowerment and Risk Across Different Levels

AI affects military affairs in two ways - it can enhance military effectiveness and also generate multiple risks. These two aspects are like the two sides of a coin, running through the various levels and domains. On the side of empowerment, AI can help automate and optimize logistics, training, command, and other areas. In direct combat and battlefield management, the emergence of unmanned platforms combined with AI is likely at least in some areas, to change the paradigm of future warfare. At the same time, however, introducing AI into the military domain brings multiple potential risks. First, it may compress decision time and increase the probability of escalation. Second, current AI systems based on neural networks and data may induce human misjudgment because of algorithmic opacity and unreliable data, potentially leading to geopolitical conflict or humanitarian tragedy. Even more subtly, AI can be used in information warfare and cognitive warfare, generating massive quantities of false audio, video, or news content to disrupt the rival's domestic public opinion and decision-making. In a context in which China and the United States are highly alert to each other's strategic intentions, this kind of information-level manipulation may impose additional shocks on strategic stability. Third, the spiral of competition driven by the intelligentization of armaments is also likely to intensify. The artificial intelligence race has clearly become a new driver and new form of military-security competition between the two countries. In the absence of mutual trust and effective communication channels, both sides fear falling behind the other, and therefore accelerate investment in armaments and defense measures. Although senior U.S. military leaders have tried to send signals of restraint to avoid an uncontrolled arms race, and China has repeatedly emphasized that it does not seek an AI arms race with other countries, the objective reality is that neither side is willing, or likely, to appear weak in the military application of artificial intelligence. New deployments by one side are often perceived by the other as pressure, prompting it to devote equivalent or greater resources to catching up. This dynamic could trap the two countries in a situation in which they "win the race but lose security." This predicament is not simply a "security dilemma." Its potential negative effects are even greater, because neither side is motivated merely by a desire to preserve the status quo; rather, both seek to use new technological possibilities to produce changes advantageous to themselves.

More specifically, the military application of AI by the two countries has different effects from the tactical level up to the strategic level. At the tactical and technical level, AI has already been widely integrated into the weapons, support platforms, and systems of both countries. Its advantages are directly reflected in agility and precision at the micro level of the battlefield management, as well as in force development and management beyond the battlefield - for example, more real-time intelligence processing, more autonomous unmanned systems, and the faster closure of the chain linking battlefield awareness and firepower, all of which can improve sensing, decision-making, and strike efficiency, as well as that of logistical

management and personnel training. This constitutes one of the most direct and intense arenas of China-U.S. competition. Both sides hope that AI will make their soldiers and weapons “smarter,” faster, and more effective than those of the other side. At the same time, the scope of these applications remains relatively limited. Whether in terms of empowerment or the risks of loss of control, there remains the possibility of spillover effects and escalation.

At the level of operations and theater, the effects of AI begin to grow more pronounced, and its influence on the shaping of the China-U.S. military posture becomes more significant. Both sides are exploring the use of AI to optimize command and control and battlefield management - for example, by using intelligent algorithms to integrate multi-source intelligence and assist commanders in decision-making, thereby creating “joint” and “all-domain” capabilities for perception and rapid response at the level of larger-scale operations, and enabling cross-service, cross-regional, and cross-domain information fusion and firepower allocation. At this level, the risks of AI applications for China-U.S. military relations also rise. Artificial intelligence-enabled rapid decision-making and joint strike capabilities mean that the offensive-defensive interactions between the two sides in potential confrontation becomes more intense, and the likelihood of loss of control and escalation increases in sensitive domains and regions. In the absence of effective mechanisms for communication and crisis management, a high-speed intelligentized battlefield could exceed the control of human commanders or “seduce” them and cause escalation from an isolated incident rapidly into a military confrontation that is difficult to contain. AI is not only a “force multiplier”; it may also become a “risk amplifier.”

At the strategic level, the effects of AI on China-U.S. military relations become still deeper and more complex, bearing directly on strategic stability, nuclear deterrence stability, and the future direction of the global security order. First, AI may undermine strategic deterrence and stability. The development of intelligentized reconnaissance and strike technologies increases the possibility that the stronger side could detect the other side’s strategic deterrent forces and carry out effective strikes against them. The prospect of depriving the other side of a second-strike nuclear capability becomes both a possibility and a temptation. Meanwhile, if the relatively weaker side judges that the survivability of its own second-strike capability is in doubt, it could in theory face a “use it or lose it” dilemma. Either mechanism would weaken the deterrent balance between the two sides. Mechanisms such as principles like "no first use of nuclear weapons" and initiatives like "mutual no-first-use" are needed to further ensure stability. In addition, the involvement of AI in strategic decision-making may lead to escalation through miscalculation. In a highly tense strategic context, if decision-makers rely excessively on information and recommendations provided by AI or AI-supported organizations or personnel, while those algorithms are biased or opaque, the training data on which they rely are flawed, or malicious manipulation, deepfakes, and cyberattacks occur, then the risk of strategic misjudgment will rise sharply. For these reasons, both China and the United States remain wary of AI at the strategic

level, including emphasizing that nuclear weapons must remain absolutely secure, reliable, and under human control. For both countries, however, the most complex challenge is that AI's involvement at the strategic level is not simply a matter of "controlling the button" or "replacing decision-making." Rather, AI is broadly "embedded" in various ways and at multiple nodes within complex human-machine loops. What major powers such as China and the United States must manage is not merely a few AI-enabled strategic tools, but vast and complex systems. In short, in the strategic domain, artificial intelligence is seen by both countries as a key commanding height in the struggle for future strategic advantage, yet because it may also shake strategic stability, it should generate concern on both sides about the serious negative consequences of not only the loss of control but beyond that. The former drives competition, while the latter creates the basis for cooperation in risk management.

An Interwoven Pattern of Competition and Cooperation

The discussion above suggests that China-U.S. military AI relations currently display an interwoven pattern in which competition and cooperation coexist. On the one hand, competition is intense. Whether in frontier technological research and development, tactical applications, operational concept innovation, or systemic modernization and adaptation, both China and the United States regard the other as the principal competitor. Both view AI as key to securing future military advantage, and both are investing heavily, concentrating resources, and striving to stay ahead in technological breakthroughs and military applications. On the question of military AI advantage, neither side is willing or likely to make concessions on its own initiative, producing a classic zero-sum dynamic.

On the other hand, the need for cooperation and risk management is equally evident. As AI continues to penetrate the strategic domain, both countries have incentives to avoid drifting into a dangerous condition in which security guardrails are absent or ineffective. In recent years, China and the United States have both begun exploring the possibility of dialogue and communication on AI risk management at the track I and track II levels. And within both countries, there have been efforts on disciplining the riskiest applications.

On the U.S. side, restraint has been institutionalized through United States Department of Defense Directive 3000.09 on autonomy in weapon systems. It requires that autonomous and semi-autonomous weapon systems be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force, and it subjects certain autonomous weapons to senior review before formal development and before fielding. Washington's Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy calls for military AI to be used responsibly, in accordance with international law, and with safeguards that reduce unintended bias, accidents, and

unintended escalation.

China's recent attempts have taken a similar but somewhat more explicitly regulatory form. Beijing has repeatedly argued that states, especially major powers, should approach the military development and use of artificial intelligence with prudence and responsibility, keep such systems safe, reliable, controllable, and under human control, and avoid misuse, abuse, and uncontrolled proliferation. China has also pushed these ideas in multilateral forums, advocating the formulation of broadly accepted rules for military artificial intelligence governance under the framework of the United Nations. It has issued the Global Artificial Intelligence Governance Initiative and its Position Paper on Regulating Military Applications of Artificial Intelligence, explicitly arguing that all countries, especially major powers, should approach the development and use of military artificial intelligence with prudence and responsibility. China has emphasized that governance of military AI should be guided by the idea of a community with a shared future for mankind, adhere to the principle of balancing development and security, and prevent an arms race.

These positions are not fundamentally at odds, in spirit, with the "responsible artificial intelligence" principles advocated by the United States; both reflect concern for artificial intelligence safety and ethics. President Xi Jinping and President Biden explicitly emphasized in the two leaders' statement at the Asia-Pacific Economic Cooperation meeting in December 2024 that humans, not artificial intelligence or machines, must retain control over nuclear weapons. This has laid an important political foundation for advancing future cooperation on relevant risk-control efforts.

There remains room for further China-U.S. coordination and cooperation on certain risk-management issues - for example, establishing communication hotlines and preventive mechanisms to avoid AI-driven crises of miscalculation, and perhaps even, in the future, limited and localized sharing of testing and validation practices and algorithmic safety-related information in order to strengthen limited two-way transparency and reduce the risk of major misjudgment. As AI technologies continue to develop, there is a real need for coordination and even cooperation between China and the United States in managing and governing risks of loss of control in military settings, human misjudgment and reckless action, and cooperation on ethical norms. It should also be noted, however, that rational recognition of common risks is not by itself sufficient to fully offset or eliminate the negative cycle of mutually reinforcing actions driven by suspicion.

This coexistence of competition and cooperation poses new challenges both for the military-security relationship between the two countries and for international security governance: how to avoid complete loss of control and unnecessary confrontation under conditions in which competition is both necessary and inevitable, and how to translate limited cooperative understandings into substantive risk management. For China and the United States, this is not only an issue between the two militaries. It is also a problem of coordination

across multiple domains and departments, spanning the entire life cycle of AI development and the full network of application and deployment. It is not only a matter of high politics and frontline policy, but also one for Track II diplomacy and academic and scientific exchange and cooperation between the two countries.

Shared Responsibility and the Outlook for Governance

As the leading countries in artificial intelligence and military technologies, the competitive and cooperative relationship between China and the United States in military AI will have a profound impact on international security in the twenty-first century. This concerns not only the strategic interests of the two countries themselves, but also peace and stability in the region and the world at large. Historical experience shows that the militarization of new technologies often precedes the establishment of corresponding rules. Major powers therefore have a responsibility to think ahead and lead in shaping governance frameworks. As artificial intelligence is integrated into the military sphere with unprecedented speed, breadth, and depth, neither China nor the United States can remain untouched, nor can either respond effectively to the structural security challenges posed by artificial intelligence through unilateral action alone. As China's white paper on arms control points out, outer space, cyberspace, artificial intelligence, and other emerging domains are new frontiers of global governance that require all countries, especially major powers, to participate jointly in formulating governance frameworks based on broad consensus.

This means that China and the United States should move beyond zero-sum mentality and engage in necessary dialogue, coordination, and even cooperation. First, the two countries should establish high-level dialogue and communication mechanisms. Drawing on the experience of the United States and the Soviet Union in establishing hotlines and signing agreements in the field of nuclear arms control, China and the United States could explore setting up regular communication channels on military applications of AI, discussing the signing of codes of conduct or the establishment of crisis communication mechanisms, and communicating promptly when AI-related accidents or anomalies occur in order to prevent escalation. It's worth noting that, such mechanisms, if possible and feasible, should narrow the mission: not "solve the crisis," but "verify whether a dangerous AI-related anomaly is real, local, degraded, spoofed, or spreading." For instance, the channel should be optimized for anomaly clarification, and incident scoping, not to settle blame in real time, but to prevent spirals driven by uncertainty. It should be a minimal template for what can be exchanged quickly during an AI-linked incident, such as time window, affected sector, preliminary confidence in attribution, indicators of potential spoofing, and steps taken to contain spread.

Second, China and the United States should play constructive roles in multilateral settings by supporting discussion, and where possible even the formulation, of international rules and standards through the United Nations, the Permanent Five, Permanent Five Plus, or other effective platforms and frameworks. The

starting point for such efforts is often low, the political resistance considerable, and the costs of coordination high. Yet even with something as minimal as jointly supporting the principle such as AI-enabled weapons and military systems should be deployed and used only under certain conditions and in certain ways, and even if such a principle lacks complete operational benchmarks for implementation, China and the United States still bear a responsibility - given the global diffusion risks of the AI race - to work together to initiate progress on the relevant agenda.

Third, China and the United States should encourage the continued strengthening of Track II exchanges and cooperation beyond official channels, and should permit and promote joint research by academic, think tank and broader epistemic communities in both countries, the development of policy initiatives, and influence on the policy cycle. They should work to build a foundational body of knowledge on the impact of artificial intelligence on military security and strategic stability, and produce practical policy recommendations. In fact, experts on both sides have already conducted extensive Track II dialogues through various platforms, and these have had tangible policy impact. The next step should be to further break through bureaucratic barriers and effectively incorporate the resulting ideas and findings into concrete military-security decision-making and capability-building processes.

In the military application of AI, China and the United States are both competitors and stakeholders facing shared risks, shouldering shared responsibilities. Both sides must avoid being drawn into a spiral of competition, while also engaging in dialogue and coordination on issues of common concern, even treating military artificial intelligence as a core issue in bilateral security dialogue on a par with nuclear weapons and strategic stability. How to strike a balance between pursuing military effectiveness, ensuring defense development and military security interests, and preventing security risks has become a common strategic challenge facing both countries and both militaries. While advancing military AI innovation and safeguarding their respective military security interests, China and the United States need to demonstrate the responsibility expected for major powers, jointly shape security frameworks and practical, context-sensitive norms and rules, and contribute to durable peace and stability in the relevant region and the world. This is not only in the long-term interests of both China and the United States, but also an inevitable requirement of acting responsibly toward the future of humankind.

(QI Haotian, Associate Professor and Deputy Head of the Department of Security Studies at the School of International Studies, Peking University (PKU))

Beyond Accusation: A Space for US-China Cooperation on AI Disinformation

DONG Ting

After years of observing US-China dialogues on artificial intelligence, one pattern is hard to miss. The agenda has actually expanded over the past two years. From military AI to frontier model risks, from biosecurity to cybersecurity, the topics under discussion are not few. Disinformation, however, has barely entered the conversation, let alone become a subject of cooperation. In existing international discussions, it usually surfaces as accusation. I want to ask whether it can move from being a topic of accusation to being a problem the two countries handle together.

Today's disinformation has a feature that is easy to overlook. Its damage often arrives before the truth is established. An unverified video, a screenshot of unknown origin, can run its full life in public discourse within a few hours. By the time the facts are pinned down, diplomatic protests have already been issued, public anger has flared, and platforms have already acted.

This kind of pressure is not new to journalism. Wait a few hours to verify, and the news cycle has moved on. Do not wait, and you risk circulating something false. But the dilemma is now spilling into diplomatic responses and policy decisions.

The current high level of mistrust between the US and China amplifies this "cannot wait" condition. A feature of low-trust environments is that anything still unclear gets sorted into the worst-case interpretation by default. An unverified video can be read, within hours, as an organized operation directed by the other government. AI does more than spread rumors faster. It collapses the time between suspicion and conviction.

This does not mean cooperation is hopeless. It can begin from a relatively modest shared recognition. If synthetic content can shape a crisis narrative within hours, either side can become a victim of that dynamic regardless of intent. The first thing the two countries need to address is stability, not truth itself.

The Cold War hotline between Washington and Moscow offers a useful reference. During the Cuban Missile Crisis, diplomatic cables between the two capitals routinely took several hours to transmit. One critical message from Khrushchev to Kennedy took close to twelve hours to receive and decode. Both sides regarded the delay as a serious risk. In June 1963, after the crisis had passed, the two governments signed a Memorandum of Understanding establishing a direct communications link. The hotline was first formally used during the 1967 Six-Day War, and again during the 1971 Indo-Pakistani War and the 1973 Yom Kippur War. It resolved no fundamental Cold War dispute. But on more than one occasion, it gave leaders on both sides a few hours of breathing room, and kept them from making decisions that could not be taken back. That same logic of buying time for

verification is what cooperation on AI disinformation could borrow today.

Beyond buying time, two other often overlooked issues belong in the same cooperative space. One is the speed gap between generation and verification. Producing a video that looks real now takes minutes. Professional forensic verification takes days or weeks. In a moment of public anger, asking a national leader to tell the public "let us wait a little longer" is politically close to impossible. Improving public media literacy will not solve this. The problem does not originate with the public.

The other issue is less visible. Future influence operations may aim less at human audiences and more at machines. As people around the world increasingly turn to large language models and AI assistants to understand US-China relations, the Taiwan Strait, or trade frictions, the low-quality content circulating online today may end up appearing as "background material" in those models' answers years from now.

The conventional response to disinformation has been takedown and labeling. But generation will keep accelerating as models improve, and removal cannot keep up. A more useful approach may be to build basic public information infrastructure. That means making a country's authoritative records and primary documents available in machine-readable form on the open internet, where human fact checkers and machines alike can find them.

What can actually be done need not be grand. Three modest steps come to my mind. First, within the existing bilateral crisis communication mechanisms, add a specific arrangement for time-sensitive information events, so that during major incidents involving AI-generated content the two sides have a few hours to consult and verify before issuing a political response. Second, push for interoperability between the two countries' technical standards on content provenance, so that the source label attached to a piece of generated imagery can still be read by the other side after it crosses the border. With that in place, no matter which side's AI tool produced a given video, the other side could trace where it came from and whether it had been altered. Third, each country should expand the volume of authoritative material it makes available in the global public information space, so that when mainstream LLMs are asked about key US-China issues, they have credible primary sources to draw on. None of these three steps requires either side to set aside existing disagreements. They share one underlying logic. When the information environment lacks basic verification capacity, the first party to lose response time is the government itself.

Structural disagreements between the US and China will not disappear, and competition in AI is likely to last. Yet even at the most strained moments, the two governments still share at least one interest. Neither has any reason to let the global public information environment become a space where no one has the time to form judgments or correct mistakes.

(DONG Ting,Fellow, Center for International Security and Strategy of Tsinghua University)

Introduction to the Artificial Intelligence Project at the Center for International Security and Strategy (CISS)

The Center for International Security and Strategy (CISS) of Tsinghua University was established on 7 November 2018. As a university-based think tank, it focuses on research in the fields of international strategy and security. Since 2019, CISS has focused on the cutting-edge developments in artificial intelligence technology and international security governance issues, establishing a dedicated expert group for its Artificial Intelligence Project to advance research on AI and international security.

Concurrently, through ongoing collaborations with the Brookings Institution in the United States and the Centre for Humanitarian Dialogue in Switzerland, CISS maintains track-two dialogues on AI and international security between China and the United States, as well as between China and Europe. CISS also continuously expands project cooperation on global AI governance with international organizations and think tanks such as the United Nations Institute for Disarmament Research (UNIDIR) and the International Committee of the Red Cross.

These efforts have produced a number of influential research outcomes and policy reports, contributing to building international consensus for advancing cooperation in the field of AI and international security. Furthermore, CISS actively undertakes research projects commissioned by national ministries including the Ministry of Foreign Affairs, Ministry of Science and Technology, and Ministry of Finance, while deepening Area Studies on AI governance to contribute to policy formulation and international cooperation.

清华大学战略与安全研究中心

CENTER FOR
INTERNATIONAL SECURITY AND STRATEGY
TSINGHUA UNIVERSITY

ciss@tsinghua.edu.cn

86-10-62771388

428A Mingli Building, Tsinghua University,
Haidian District, Beijing, China (100084)

http://ciss.tsinghua.edu.cn

WeChat Website Contact Us