

How to cite this article:
Ameriyan, H. R., Mirkooshesh, A. H., & Razani, E. (2027). Speed–Safety Asymmetry in the U.S.–China Military Artificial Intelligence Competition: Implications for Strategic Stability and the Growing Risk of Instantaneous War. *Journal of Historical Research, Law and Policy, 5*(9), 1-22. https://doi.org/10.61838/jhrlp.491

Article history:
Original Research

Dates:
Submission Date: 20 July 2026
Revision Date: 23 October 2026
Acceptance Date: 29 October 2026
First Pubication Date: 08 November 2026
Final Publication Date: 01 September 2027

Speed–Safety Asymmetry in the U.S.–China Military Artificial Intelligence Competition: Implications for Strategic Stability and the Growing Risk of Instantaneous War

1. Hamid Reza Ameriyan: Department of Political Science and International Relations, Sha.C., Islamic Azad University, Shahrood, Iran
2. Amir Hooshang Mirkooshesh*: Department of Political Science and International Relations, Sha.C., Islamic Azad University, Shahrood, Iran
3. Ehsan Razani: Department of Political Science and International Relations, Sha.C., Islamic Azad University, Shahrood, Iran

*corresponding author's email: am.mirkooshesh@iau.ac.ir

ABSTRACT

The competition between China and the United States in the field of military artificial intelligence has emerged in recent years as one of the most significant sources of tension in relations between the two major powers. What distinguishes this competition is the profound divergence in the priorities of the two countries: the United States continues to emphasize maintaining meaningful human control and ensuring system safety, whereas China places greater priority on rapid deployment and higher levels of machine autonomy. This study focuses specifically on this asymmetry and seeks to examine its implications for strategic stability. The research employs a comparative–analytical approach and draws on game theory and the concept of the technological security dilemma as its principal analytical frameworks. The findings indicate that this asymmetry increases time pressure on U.S. decision-makers, heightens the likelihood of algorithmic misperception, and weakens traditional arms-control mechanisms, thereby substantially increasing the risk of instantaneous war. Based on the findings, establishing dedicated communication channels for managing algorithmic crises, reaching agreement on minimum red lines—particularly in the nuclear domain—and strengthening selective transparency could mitigate this risk to some extent.

Keywords: *military artificial intelligence, speed–safety asymmetry, strategic stability, instantaneous war, United States, China, autonomous weapon systems*

Introduction

Over the past two decades, artificial intelligence has moved beyond the status of an emerging technology and has become one of the most consequential determinants of power and security in the international system. This technology is no longer merely a supplementary tool; rather, as a "general-purpose technology," it is redefining the nature of warfare, the logic of deterrence, and the mechanisms of strategic decision-making. The integration of machine-learning algorithms and autonomous systems into command, control, intelligence, and the application of lethal force has compressed decision-making time to seconds and microseconds and has confronted classical

© 2027 the authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

concepts such as the OODA loop, meaningful human control, and strategic stability with unprecedented challenges. Competition between the United States and China has likewise moved beyond a conventional arms race and evolved into a contest over algorithmic superiority and the speed with which innovation can be converted into operational capability.

At the heart of this competition lies a structural asymmetry. Drawing on its legal and institutional traditions, the United States emphasizes the preservation of "meaningful human control." U.S. Department of Defense Directive 3000.09, as updated in 2023, provides that autonomous systems should be designed in such a way that commanders can exercise appropriate levels of human judgment over decisions relating to the use of lethal force. This cautious approach strengthens legitimacy and accountability, but it also slows the pace of development. By contrast, China, within the framework of its doctrine of "intelligentized warfare" and its military-civil fusion strategy, places greater priority on rapid deployment and higher levels of machine autonomy. From Beijing's perspective, in environments such as the Taiwan Strait, insistence on continuous human supervision can become a systemic bottleneck. This difference between an "emphasis on safety and human control" and an "emphasis on speed and autonomy" constitutes the core of the "speed–safety asymmetry."

The importance of examining this issue arises from several considerations. First, the existing literature has largely focused either on the technical and ethical dimensions of autonomous systems or on broader geopolitical competition, while the implications of this doctrinal asymmetry for strategic stability and the risk of "instantaneous war" have received comparatively limited analysis. Second, the compression of decision-making time and the possibility of direct interaction between hostile algorithms have brought the risk of unintended escalation closer to the operational level. Third, traditional arms-control regimes have become increasingly ineffective in dealing with the software-based and largely invisible nature of algorithms, while the resulting normative vacuum further intensifies competition.

The principal research question is as follows: What implications does the asymmetry between China's prioritization of speed and the United States' prioritization of safety have for bilateral strategic stability and the risk of instantaneous war? The subsidiary questions examine the institutional roots of this asymmetry, the pathways through which it increases the risks of miscalculation and algorithmic misperception, and its implications for arms-control regimes.

The principal objective of this article is to provide a systematic analysis of this asymmetry and assess its implications for strategic stability. The secondary objectives include explaining its underlying causes, identifying mechanisms of algorithmic escalation, and proposing policy measures to reduce the risk of instantaneous war.

Theoretical Framework and Literature Review

Game Theory and the Technological Security Dilemma

Game theory, as one of the most powerful analytical tools in strategic studies and international relations, provides a mathematical and logical framework for understanding interactions among rational actors in which the final outcome is not merely a function of one actor's decision but depends on the reciprocal choices of others. (1) The modern foundations of this theory were first formulated by John von Neumann and Oskar Morgenstern and were later developed through John Nash's introduction of the concept of "Nash equilibrium," which transformed game theory into a central analytical instrument for examining state behavior under conditions of competition and

uncertainty. (2) A Nash equilibrium is a situation in which no actor can increase its utility by unilaterally changing its strategy, provided that the strategies of the other actors remain unchanged. This concept has been widely applied to the analysis of arms races, deterrence, and security dilemmas and enables the modeling of state behavior in situations where outcomes depend on reciprocal decisions.

In the field of technological and arms competition, game theory constitutes an indispensable conceptual tool for understanding the behavioral logic of major powers. When two or more actors operate under conditions of structural mistrust and fear of relative decline, even if both recognize the negative consequences of excessive acceleration, the incentive structure of the game may drive them toward choices that produce collectively harmful outcomes. (3) This dynamic is particularly visible in the competition between China and the United States over the development and deployment of military artificial intelligence—a competition concerned not merely with the number of systems, but also with the speed of algorithmic decision-making, levels of autonomy, and the prioritization of safety relative to operational effectiveness.

The Prisoner's Dilemma in the Military AI Arms Competition

One of the most classical and widely applied models of game theory that effectively explains the dynamics of the military AI arms competition is the "prisoner's dilemma." In its standard formulation, two actors face two options: cooperation or defection. If both parties choose cooperation, a collectively desirable outcome is achieved and relative security is preserved. However, because of structural mistrust, the absence of credible commitment mechanisms, and fear of exploitation by the other side, the dominant strategy for both actors becomes defection. The resulting Nash equilibrium is suboptimal for both parties and reduces collective welfare. (4)

In his classical formulation of the "security dilemma," Robert Jervis demonstrated that measures adopted by one state to increase its own security—including the development of new weapons—can inherently reduce the security of other states and compel them to respond in kind, thereby ultimately weakening the security of the system as a whole. (3) This logic is reproduced with even greater intensity in the military AI arms competition. Both China and the United States confront two strategic options. The first is "cooperation," understood as cautious, safe, and controlled development of military AI accompanied by the preservation of meaningful levels of human control and compliance with testing and validation standards. The second is "defection," meaning the acceleration of the development and deployment of autonomous systems by prioritizing speed, relatively reducing safety considerations, and accepting higher levels of machine autonomy. (5, 6)

The logic of the prisoner's dilemma suggests that, because of mutual fears of technological inferiority, strategic surprise during crises, or the loss of decision-making superiority in the opening hours of conflict, the dominant strategy for both parties will be defection. Consequently, the Nash equilibrium emerges at a point where both countries have deployed systems with higher levels of autonomy and lower control standards. Such a situation not only reduces collective security but also significantly increases the risks of accidental war, unintended crisis escalation, and algorithmic incidents. (5, 7) What appears at the individual level to be a rational measure for preserving national security can therefore produce reduced security for both sides and strategic instability at the systemic level.

Suboptimal Nash Equilibrium and the Race to the Bottom in Standards

The Nash equilibrium arising from the prisoner's dilemma in military AI competition is suboptimal in nature. In such an equilibrium, both actors prefer to lower safety standards and requirements for human control in order to avoid relative technological disadvantage, even though both recognize that mutual adherence to these standards would produce a better outcome for all parties. This dynamic creates the conditions for what contemporary security studies describe as a "race to the bottom" or reciprocal erosion of standards. (8, 9)

In this process of reciprocal erosion, each side gradually reduces or circumvents algorithmic safety standards, meaningful human-control requirements, and even certain testing and validation protocols in order to avoid relative technological inferiority or strategic surprise. Fear that the rival may achieve operational capability first or enjoy a speed advantage during a crisis weakens incentives to comply with precautionary constraints. The result is the gradual erosion of control norms and an increased probability of unintended algorithmic incidents. (6, 7) In such an environment, behavior that appears rational at the individual level produces greater systemic risk and increased fragility of strategic stability at the collective level.

This confrontation becomes particularly acute when doctrinal asymmetry exists between the two powers. When one side gives priority to greater speed and autonomy while the other emphasizes human control and safety, pressure on the more cautious actor to lower its standards increases. In real-world crises, this pressure can generate a strategic dilemma: either accept the risk of falling behind in the decision cycle or temporarily or permanently reduce control constraints in order to match the rival's speed. Both options weaken strategic stability and increase the risk of unintended escalation. (5, 9)

Thus, game theory and the concept of the technological security dilemma provide an appropriate analytical framework for understanding the behavioral logic of China and the United States in military AI competition. This framework demonstrates that even where both parties recognize the negative consequences of excessive acceleration and declining safety standards, the incentive structure of the game can drive them toward a suboptimal Nash equilibrium and reciprocal erosion of standards. Such an equilibrium not only reduces collective security but also significantly increases the risk of instantaneous war and weakens strategic stability, thereby highlighting the need to consider confidence-building mechanisms and minimum control regimes. (3, 4, 9)

The Concept of Strategic Stability and Its Transformation in the Age of Artificial Intelligence

Since the end of the Second World War and the beginning of the nuclear age, the concept of "strategic stability" has become one of the most central ideas in security studies. In the classical deterrence literature, strategic stability refers to a condition in which neither side possesses the incentive or effective capability to launch a successful preemptive strike, crises remain manageable at lower levels of escalation, and deterrence mechanisms prevent tensions from escalating into full-scale conflict. (10)

During the Cold War, this concept rested on two principal foundations: "crisis stability," meaning the absence of incentives for a first strike even at the peak of tensions, and "arms-race stability," meaning the absence of incentives for unrestricted weapons accumulation. Schelling and Halperin emphasized that stability exists when parties believe that a preemptive attack could result in mutual destruction.

The advent of artificial intelligence has transformed these foundations. Decision-making time has become highly compressed, transparency has declined, and the possibility of unintended crisis escalation through autonomous

systems has increased. Strategic stability is therefore no longer defined solely in terms of the number of warheads or second-strike capability, but increasingly depends on the speed of data processing, levels of system autonomy, and the ability of human operators to intervene in a timely manner. (6, 7)

In the age of artificial intelligence, both dimensions of stability face significant challenges. Autonomous systems have reduced reaction times to fractions of a second and constrained opportunities for human intervention. The software-based and largely invisible nature of AI capabilities has also made transparency and verification more difficult. As a result, crisis stability has become more fragile, while competition for algorithmic superiority has weakened arms-race stability. (5, 6)

One of the most profound effects is the compression of decision-making time. The traditional OODA cycle, which depended on human cognitive capacity, can now be reduced to seconds and microseconds. This speed challenges established concepts of deterrence and crisis management. (6) The concept of "hyperwar," introduced by Allen and Husain, refers to conflict in which the speed of processing and combat action exceeds human capacity, turning the human operator into a systemic bottleneck. (11)

This temporal compression has several consequences: opportunities for diplomacy decline, the risks of algorithmic misperception and unintended escalation increase, and pressure grows to delegate greater autonomy to machines. Consequently, the classical foundations of strategic stability—transparency, sufficient time for human decision-making, and confidence in second-strike capability—face structural challenges. Preserving stability under such conditions requires reconsideration of human control, confidence-building regimes, and the deliberate management of algorithmic speed.

The Concept of "Instantaneous War"

One of the most concerning consequences of integrating artificial intelligence into military systems is the possibility of conflicts unfolding at speeds beyond the capacity for deliberate human intervention. This phenomenon is increasingly discussed in contemporary strategic literature as "instantaneous war." Instantaneous war refers to conditions in which hostile algorithmic or semi-autonomous systems become trapped in an escalating cycle of action and reaction capable of transforming a manageable tension into large-scale military confrontation within fractions of a minute or even seconds. (5)

This concept arises directly from the transformation that artificial intelligence has introduced into the military decision-making loop. When system reaction times shift from the human scale to the machine scale, the possibility of managing a crisis incrementally is substantially reduced. In such an environment, even a minor misperception, sensor error, or hostile interpretation of an adversary's action can trigger a chain of automated responses that acquires strategic significance before human actors have an opportunity to assess the situation. (6, 7)

Comparison with Financial-Market Flash Crashes

Paul Scharre conceptualizes instantaneous war by drawing an analogy with the phenomenon of "flash crashes" in financial markets. On May 6, 2010, automated trading algorithms caused severe declines in U.S. stock-market indices within minutes; without coordinated human intervention, the consequences could have been considerably more severe. (5) What occurred was an interaction among algorithms, each of which acted according to its own optimization logic but collectively produced systemic breakdown.

Scharre argues that a similar logic could be reproduced on an algorithmic battlefield. When autonomous systems from opposing sides interact without effective human supervision, a form of “flash war” may emerge, in which the cycle of action and reaction proceeds at the speed of data transmission and the crisis moves beyond control before human operators fully understand its dimensions. (5) The fundamental difference is that in the military domain, the consequences are not financial losses but human casualties, infrastructure destruction, and the possibility of escalation to higher levels, including the nuclear domain.

This comparison demonstrates that the danger of instantaneous war does not necessarily arise from hostile intent; rather, it may result from the complex and unpredictable interaction of systems that each operate “rationally” in isolation. (5, 7)

The Risk of Algorithmic Interaction Without Human Intervention

The principal danger of instantaneous war lies in direct, unmediated interaction among hostile algorithms. As levels of autonomy increase and humans are removed from or marginalized within tactical decision-making loops, opportunities for corrective intervention decline substantially. If one side’s algorithm interprets the other side’s action as hostile, a chain reaction can unfold at the speed of data transmission, eliminating the opportunity for human assessment. (5, 6)

This danger is particularly pronounced in environments such as the Taiwan Strait or the South China Sea. An aerial interception, naval maneuver, or even a simple technical malfunction could be interpreted by the opposing side’s automated recognition systems as definitive evidence of aggression, thereby closing the temporal window for political consultation and emergency communication channels. (6, 7)

The phenomenon of “automation bias” further magnifies this danger. When operators observe successful system performance over extended periods, they may develop excessive trust in the system and lose either the capacity or motivation to challenge machine-generated outputs during critical moments. (5) Consequently, even in systems nominally characterized by a “human-on-the-loop” architecture, effective intervention can be weakened.

Thus, instantaneous war represents one of the most profound challenges to strategic stability in the age of artificial intelligence. It is not merely the product of hostile intent but a logical consequence of compressed decision-making time, increasing machine autonomy, and diminishing opportunities for human intervention. Understanding its mechanisms and designing effective arrangements to preserve meaningful human control have therefore become strategic imperatives.

Literature Review

The literature concerning the role of artificial intelligence in military affairs and its implications for strategic stability has expanded considerably over the past decade. Nevertheless, a substantial portion of this literature has focused either on the technical and ethical dimensions of autonomous systems or on U.S.–China competition at the broader level of geopolitics and technological innovation. A critical review of key works indicates that although significant advances have been made in understanding this domain, systematic analysis of the doctrinal asymmetry between the two powers and its relationship to the risk of instantaneous war remains fragmented.

One of the foundational works in this field is Paul Scharre’s *Army of None*. Scharre carefully conceptualizes autonomy in weapon systems and distinguishes among different levels of human involvement in the decision-making loop. (5) He also develops the concept of “instantaneous war” by drawing on the phenomenon of financial-

market flash crashes and clearly demonstrates the risks associated with interactions among hostile algorithms in the absence of effective human intervention. (5) Despite its analytical richness, the book focuses primarily on the technical, ethical, and operational dimensions of autonomous systems, while in-depth comparative analysis of Chinese and U.S. doctrines and institutional priorities remains relatively limited.

Michael Horowitz, in a series of studies and particularly in his article “When Speed Kills,” emphasizes the implications of decision-time compression for deterrence and strategic stability. He argues that machine-driven speed can challenge traditional concepts of crisis management and increase the risk of unintended escalation. (6) Horowitz also highlights the entanglement of conventional command-and-control systems with nuclear systems and draws attention to the risk that tensions could spread from the conventional to the strategic level. However, rather than concentrating on doctrinal asymmetry between China and the United States, his analysis focuses more broadly on the implications of algorithmic speed for stability.

Kenneth Payne, in *I, Warbot*, examines in greater detail the concept of hyperwar and the gradual removal of human agency from decision-making chains. Payne demonstrates that artificial intelligence compresses the temporal dimension of warfare and turns humans into a systemic bottleneck. (7) He also addresses the cognitive and psychological challenges arising from increased reliance on algorithms, including automation bias. Despite the theoretical precision of Payne’s work, his analysis is oriented more toward general transformations of the algorithmic battlefield than toward a systematic comparison of the approaches adopted by the two major powers.

At the level of policy reports, analyses of U.S.–China military AI competition have emphasized the importance of algorithmic superiority for future warfare and warned that falling behind in this domain could undermine traditional U.S. military advantages. (9) Reports focusing specifically on U.S.–China competition have identified multiple pathways through which military AI may increase strategic risk and have emphasized the danger of unintended escalation. (12) RAND studies, including simulations related to Taiwan scenarios, likewise indicate that the use of autonomous systems may generate new patterns of attritional conflict and fundamentally alter commanders’ decision-making environments. (13)

In studies of Chinese doctrine, Elsa Kania was among the first researchers to formulate the concept of “battlefield singularity” and analyze the intelligentization strategy of the People’s Liberation Army. (14) Subsequent studies have likewise emphasized China’s focus on rapid deployment, military-civil fusion, and a willingness to adopt higher levels of autonomy in operational environments close to its territory. However, much of this literature has either focused on China’s technical capabilities or examined their implications at the broad geopolitical level, paying less systematic attention to the relationship between these priorities, the speed–safety asymmetry, and the risk of instantaneous war.

A critical review of the foregoing literature indicates that, although the field is rich and expanding, it suffers from several structural limitations. First, in-depth comparative analyses of the institutional and doctrinal approaches of China and the United States remain limited. Second, the relationship between speed–safety asymmetry, instantaneous war, and strategic stability has not been systematically conceptualized. Third, a substantial portion of the literature either focuses on one power or addresses the general implications of artificial intelligence, while paying relatively little attention to the interactive dynamics between the two sides. The present article seeks to address this gap by treating speed–safety asymmetry as the central analytical variable and conceptualizing its implications for strategic stability and the risk of instantaneous war.

Methodology

The present study employs a comparative–analytical method to examine the speed–safety asymmetry in military AI competition between the United States and China. The comparative method enables the systematic identification of similarities and differences between the two actors across institutional, doctrinal, and operational dimensions, while the analytical method provides a basis for interpreting these differences through the concepts of strategic stability, the security dilemma, and instantaneous war. The combination of these two approaches is particularly appropriate for studies of great-power competition and emerging technologies because it enables analysis to move beyond mere description of documents toward explanation of their strategic implications. (5, 6)

The choice of this method is necessary because the research problem is inherently bilateral and interactive. Speed–safety asymmetry is not a static characteristic of one country's policy but rather the product of dynamic interactions among the priorities, institutional constraints, and mutual perceptions of the two powers. Separate analysis of each side's approach, without systematic comparison, cannot fully explain the resulting destabilizing dynamics. The comparative method makes it possible to place the two approaches in direct relation to one another and identify points of friction, while the analytical dimension connects these findings to the theoretical framework of strategic stability and the risk of instantaneous war. (7)

Data Sources

The research data were collected from three principal categories of sources. The first category consists of official and doctrinal documents from the two countries. With regard to the United States, Department of Defense Directive 3000.09, as revised in 2023, was examined as the most significant legal-institutional document governing autonomy in weapon systems. This document specifies requirements for maintaining appropriate human judgment in the application of lethal force. (9) With regard to China, documents relating to the doctrine of "intelligentized warfare," the military-civil fusion strategy, and the priorities of the People's Liberation Army constitute the basis of the analysis. Although these sources are less transparent, they reveal Beijing's orientation toward prioritizing speed and autonomy. (7, 14)

The second category consists of reports by reputable think tanks. Reports by the Center for a New American Security, RAND, and the Stockholm International Peace Research Institute are valuable because of their reliance on empirical data and technical-strategic analysis. The CNAS report conceptualizes pathways through which strategic risk may increase, (12) RAND studies examine the operational implications of autonomous systems in a Taiwan scenario, (13) and SIPRI reports provide valuable comparative data on autonomy. (15)

The third category consists of academic and specialized literature, including works by researchers such as Scharre, Horowitz, Payne, and Kania, which provide the conceptual framework and make it possible to evaluate the findings against the existing literature. (5-7, 14)

Comparative Criteria

Four principal criteria were selected to organize the comparative analysis:

1. **Level of human control:** the extent to which decision-making authority is delegated to machines and the position of humans within the control loop—human-in-the-loop, human-on-the-loop, or human-out-of-the-

loop. This criterion is derived directly from Scharre's conceptualization and relevant U.S. policy frameworks. (5, 9)

2. **Deployment speed:** the temporal distance between laboratory innovation and operational field capability, as well as each side's willingness to accept risk in testing and deployment. This criterion is essential for understanding the asymmetry between the more cautious U.S. approach and China's acceleration-oriented approach. (12)
3. **Doctrinal priorities:** overarching military objectives, perceptions of the character of future warfare, and the relative weight assigned to factors such as tactical surprise, network resilience, and human control. (7, 14)
4. **Implications for strategic stability:** the effects of the interaction among the preceding three criteria on crisis stability, the risk of unintended escalation, and the possibility of instantaneous war. This criterion connects the descriptive findings to the central research question. (5, 6)

The combination of these four criteria makes it possible to move beyond the description of documents and practices toward an explanation of the structural consequences of speed–safety asymmetry.

Comparative Analysis of U.S. and Chinese Approaches

The U.S. Approach: Prioritizing Safety and Meaningful Human Control

The U.S. approach to the development and employment of military artificial intelligence has been shaped above all by the country's legal, ethical, and institutional traditions. Rather than focusing exclusively on achieving superior speed or maximum autonomy, official U.S. policy emphasizes maintaining meaningful levels of human control and judgment in decisions concerning the use of lethal force. This priority reflects deep concerns regarding accountability, domestic and international legitimacy, and the risk of unintended incidents and is reflected in official documents, procurement processes, and the organizational culture of the Department of Defense. (5, 9)

Although this cautious approach constitutes a normative and legal strength, in practice it is associated with structural constraints and can slow the process of translating laboratory innovations into operational capabilities. The principal dimensions of this approach are examined below under three headings.

Department of Defense Directive 3000.09 (2023 Version)

The most significant legal-institutional document governing autonomy in U.S. weapon systems is Directive 3000.09, which was updated in January 2023. The directive defines an autonomous weapon system as one that, once activated, can select and engage targets without further intervention by a human operator. (9) At the same time, it emphasizes that the design, development, and employment of such systems must allow commanders to exercise appropriate levels of human judgment over decisions concerning lethal force.

The 2023 version imposes more stringent requirements than earlier iterations regarding testing, validation, and hierarchical oversight. The deployment of systems with higher levels of autonomy requires multilayered approval from senior civilian authorities within the Pentagon. This framework clearly prioritizes safety, reliability, and accountability and seeks to prevent the complete delegation of life-and-death authority to algorithms. (5, 9) At the same time, these multilayered requirements make development slower and more costly and may become an operational disadvantage relative to an adversary less constrained by comparable limitations.

Human-in-the-Loop and Human-on-the-Loop Models

The U.S. approach is based on distinctions among different levels of human involvement in the control loop. In the “human-in-the-loop” model, a system may independently search for, track, and identify targets, but the final decision to employ lethal force remains exclusively with a human operator. At this level, the machine functions as an instrument for augmenting human cognitive capabilities, and no lethal action occurs until a human issues the command to engage. (5)

In the “human-on-the-loop” model, the system possesses greater operational independence and may proceed through the entire targeting cycle, but remains under real-time human supervision. The human functions as a safety monitor and can interrupt or modify the process if an error is detected or circumstances change. (5, 9) Missile-defense systems such as Aegis, in certain operational modes, provide examples of this level of autonomy.

The United States has officially avoided a complete transition to the “human-out-of-the-loop” model in systems associated with lethal force and continues to insist on preserving minimum meaningful levels of human control. This position is rooted both in legal and ethical considerations and in concerns about the strategic consequences of fully delegating authority to machines. (6, 7)

Institutional Barriers

Despite the relative clarity of doctrinal documents, the U.S. approach faces numerous institutional barriers that slow the operationalization of technology. The first obstacle is the complex and multilayered nature of procurement bureaucracy. The U.S. defense acquisition system, with its emphasis on competition, transparency, and accountability, devotes significant time to evaluation and approval processes. Although these procedures are desirable from a governance perspective, they lengthen the interval between laboratory innovation and field capability. (9)

The second obstacle is the well-known “valley of death,” referring to the gap between initial prototyping and mass production or operational deployment. Many innovative military AI projects are delayed or terminated at this stage because of budgetary complexities, changing priorities, or misalignment with commanders’ operational requirements. (5, 9)

The third obstacle concerns stringent legal and ethical requirements. Emphasis on individual accountability, compliance with international humanitarian law, and sensitivity to domestic public opinion have encouraged the Department of Defense to adopt a cautious approach toward higher levels of autonomy. This approach strengthens the legitimacy and social acceptability of the technology, but when competing with an actor that prioritizes speed over safety, it may contribute to relative delays in operational deployment. (6, 7)

Overall, the U.S. approach rests on the prioritization of safety, meaningful human control, and accountability. Directive 3000.09, the human-in-the-loop and human-on-the-loop models, and existing institutional barriers all reflect a structural preference for preserving human agency in lethal decision-making. However, these same characteristics slow the translation of innovation into operational capability and may create strategic pressure when confronting an adversary that places greater priority on speed.

The Chinese Approach: Prioritizing Speed and Higher Levels of Autonomy

China's approach to military artificial intelligence, in contrast to that of the United States, places a clearer priority on deployment speed, tactical surprise, and higher levels of machine autonomy. This orientation derives from Beijing's strategic perception of future warfare, the requirements of operational environments close to its own borders, and the belief that it is necessary to neutralize U.S. network-centric advantages. Three principal dimensions characterize this approach: the doctrine of "intelligentized warfare," the military-civil fusion strategy, and a tendency toward higher levels of autonomy in scenarios such as the Taiwan Strait and the South China Sea. (7, 14, 16) Whereas the United States emphasizes meaningful human control and precautionary procedures, China places greater weight on reaction speed, network resilience under conditions of communications disruption, and the capacity to achieve surprise. This doctrinal difference constitutes the central foundation of the speed–safety asymmetry and has direct implications for strategic stability.

The Doctrine of Intelligentized Warfare

The concept of "intelligentized warfare" is presented in Chinese military literature as a stage beyond informationized warfare. From this perspective, artificial intelligence is no longer merely a supporting instrument but a central driver of transformation in command, decision-making, and operational execution. (14, 17) The People's Liberation Army considers superiority in data processing, cognitive algorithms, and autonomous systems capable of neutralizing traditional U.S. military network advantages and opening pathways to victory in technologically advanced local wars.

Elsa Kania has shown that China pursues the concept of a "battlefield singularity" as a desirable future condition in which artificial intelligence surpasses human capacity in the speed and quality of decision-making. (14) Within this framework, insistence on complete and continuous human supervision during critical moments may be perceived not as a strength but as a systemic bottleneck capable of contributing to failure. (7) For this reason, the doctrine of intelligentized warfare implicitly, and at times explicitly, incorporates a willingness to grant machines greater levels of autonomy.

Military-Civil Fusion and the Emphasis on Surprise

One of China's most important mechanisms for accelerating the development of military AI is its military-civil fusion strategy. Under centralized state direction, this strategy mobilizes civilian-sector resources—including technology companies, universities, and vast domestic data resources—for defense objectives. (12, 14) Unlike the more decentralized and market-oriented U.S. model, this integration can shorten the distance between laboratory innovation and operational capability and accelerate testing and deployment.

Alongside this institutional mechanism, Chinese doctrine places considerable emphasis on tactical surprise. Beijing assumes that in environments close to its homeland—particularly scenarios involving Taiwan—the ability to strike rapidly, disrupt adversary command systems, and sustain operations under conditions of electronic warfare and satellite-communications disruption will be decisive. (16, 18) In such an environment, systems capable of continuing operations without continuous connectivity to a ground station or real-time human approval acquire greater operational value and reinforce incentives for higher levels of autonomy.

Preference for Higher Autonomy in Proximate Operational Environments

China's willingness to accept higher levels of autonomy is most visible in scenarios close to its own territory. In these environments, the density of military forces, geographical proximity, and the likelihood of severe communications disruption create conditions in which continuous human supervision may become difficult or even impossible. (7, 12) From Beijing's perspective, systems capable of continuing their missions in communications-denied environments can generate asymmetric advantages.

Existing analyses indicate that China has moved more rapidly toward higher autonomy in areas such as loitering munitions, autonomous swarming systems, and the integration of artificial intelligence into command networks. (9, 14) This does not necessarily imply complete acceptance of a human-out-of-the-loop model, but compared with the more cautious U.S. approach, it reflects a clearer prioritization of speed and machine independence.

Overall, China's approach is shaped by the doctrine of intelligentized warfare, the military-civil fusion strategy, and the demands of proximate operational environments. This approach places greater priority on deployment speed and higher autonomy than on continuous human control and therefore constitutes one of the two principal poles of the speed–safety asymmetry. The implications of this prioritization for strategic stability include greater temporal pressure on the adversary and an increased risk of unintended escalation during real-world crises.

Identification and Explanation of the Speed–Safety Asymmetry

A systematic comparison of U.S. and Chinese approaches to military artificial intelligence reveals a structural and persistent asymmetry that can be described as the "speed–safety asymmetry." This asymmetry is not a minor difference in technical priorities but reflects deeper differences in strategic culture, institutional architecture, and perceptions of the nature of future warfare. By emphasizing meaningful human control, accountability, and precautionary processes, the United States prioritizes safety and legitimacy over speed. China, drawing on its doctrine of intelligentized warfare and military-civil fusion strategy, treats deployment speed, tactical surprise, and higher levels of autonomy as decisive. (7, 9, 14) Identifying this asymmetry and its operational consequences in real crises is essential for understanding the destabilizing dynamics of great-power competition.

Systematic Comparison of the Two Approaches

From the perspective of human control, the difference is clear. In official policy, particularly the framework reflected in Directive 3000.09, the United States emphasizes the preservation of appropriate human judgment in decisions concerning lethal force and treats human-in-the-loop and human-on-the-loop arrangements as preferred models. (5, 9) Within the framework of intelligentized warfare, China is more likely to view continuous human supervision as a systemic bottleneck in complex and communications-denied environments and demonstrates greater willingness to delegate higher levels of autonomy. (7, 14)

From the perspective of deployment speed, the United States faces numerous institutional barriers, including complex bureaucratic procedures, the "valley of death," and legal-ethical requirements that lengthen the interval between laboratory innovation and field capability. (5, 9) By employing military-civil fusion and centralized state direction, China has shortened this interval and moved more rapidly in areas such as swarm systems and loitering munitions. (12, 14)

Fundamental differences are also evident in doctrinal priorities. The United States emphasizes maintaining legitimacy, accountability, and compliance with international humanitarian law and therefore subjects speed to these considerations. China, by contrast, focuses on competition among operational systems and on disrupting, paralyzing, or destroying critical elements of adversary command-and-control systems, employing autonomy and network integration in pursuit of these objectives. (6, 16, 18)

Taken together, these three dimensions—level of human control, deployment speed, and doctrinal priorities—have produced a persistent asymmetry in which one side prioritizes safety over speed while the other places comparatively greater emphasis on speed and autonomy. This asymmetry is rooted in the structural differences between the two political-military systems and is therefore unlikely to be easily eliminated.

Operational Implications in Real-World Crises

The consequences of this asymmetry become most visible in real-world crisis scenarios, particularly in crisis-prone regions such as the Taiwan Strait and the South China Sea. In these environments, dense military presence, geographical proximity, and the possibility of severe communications disruption make reaction speed a decisive variable. (7, 12)

If Chinese autonomous or semi-autonomous systems operate at greater speed and with less human supervision, the U.S. command chain may face increasing temporal, operational, and psychological pressure. U.S. commanders would either have to accept the risk of falling behind in the decision cycle or reduce their own safety protocols in order to keep pace with the adversary. (5, 6) Both options weaken strategic stability: the first increases the risk of tactical defeat, while the second facilitates a “race to the bottom” in oversight standards.

Moreover, the asymmetry may intensify mutual misperception. Washington may interpret greater Chinese system autonomy as evidence of aggressive intent, while Beijing may interpret the more cautious U.S. approach as evidence of slowness or weakness. (9, 12) Such pessimistic interpretations increase incentives to prepare on the basis of worst-case scenarios and lower the threshold for crisis escalation.

Ultimately, this asymmetry intensifies the danger of instantaneous war. When one side operates at machine speed while the other experiences delays associated with human control, the likelihood increases that escalating cycles of action and reaction will develop beyond the capacity for timely human intervention. (5, 6) Under such conditions, even a minor misperception or algorithmic error could transform a manageable crisis into large-scale conflict.

Thus, the speed–safety asymmetry is not an abstract characteristic but an operational variable with tangible consequences for strategic stability. It is rooted in institutional and doctrinal differences and, in real-world crises, can generate temporal pressure, increase misperception, and facilitate pathways toward unintended escalation.

Implications of the Asymmetry for Strategic Stability

Increasing Time Pressure and the U.S. Strategic Dilemma

One of the most direct and dangerous consequences of the speed–safety asymmetry is the increasing temporal pressure imposed on the U.S. command-and-control chain during crises. When China prioritizes deployment speed and higher autonomy while the United States insists on maintaining meaningful human control, the more cautious actor confronts a difficult dilemma at critical moments: either surrender the initiative and accept the risk of falling

behind in the decision cycle or reduce its standards of human control and safety in order to match the adversary's speed. (5-7) Both options weaken strategic stability and increase the risk of unintended escalation.

This temporal pressure is rooted in the severe compression of the decision-making loop. In algorithmic warfare, autonomous or semi-autonomous systems may complete the cycle of observation, orientation, decision, and action in fractions of a second. If one side's systems operate with less human supervision and greater speed, the opposing side, which considers itself bound by requirements for human approval, is structurally placed at a temporal disadvantage. (6, 7) At flashpoints such as the Taiwan Strait or the South China Sea, where military density and the likelihood of communications disruption are high, this temporal asymmetry can rapidly spread from the tactical to the strategic level. (12)

The U.S. dilemma can be conceptualized along two pathways. The first involves full adherence to existing protocols and acceptance of the risk of surrendering the initiative. Under this scenario, U.S. commanders may lose the opportunity to respond effectively to faster Chinese systems and could concede the battlespace during the opening hours or even minutes of a crisis. Although this option is consistent with the legal and ethical requirements underlying Directive 3000.09, it could lead to tactical defeat or the loss of critical positions and create political pressure to revise existing constraints. (6, 9)

The second pathway involves temporarily or permanently reducing human-control standards in order to recover speed. Under conditions of intense temporal and psychological pressure, commanders may be pushed toward delegating higher levels of autonomy. Even if such delegation is limited and context-specific, it could establish a dangerous precedent and facilitate a "race to the bottom" in oversight standards. (5, 12) Automation bias further intensifies this risk because operators who have observed successful system performance may become more willing to accept machine recommendations during crises. (5, 7)

Both pathways weaken strategic stability. Surrendering the initiative increases the risk of defeat and escalation from a position of weakness. Reducing control standards increases the probability of algorithmic errors, misperception, and escalating cycles of action and reaction. In both cases, the temporal window available for political intervention and crisis diplomacy becomes narrower, increasing the danger that a manageable tension will develop into broader conflict. (5, 6)

Moreover, the adversary's awareness of this dilemma may itself become destabilizing. If Beijing calculates that temporal pressure could push Washington toward reducing its standards or accepting greater risks, incentives may increase to create rapid faits accomplis or test U.S. red lines. Such dynamics intensify the logic of the security dilemma and make crisis stability more fragile. (6, 12)

Overall, the increase in temporal pressure arising from speed–safety asymmetry places the United States in a difficult strategic dilemma. Both surrendering the initiative and reducing standards of human control entail substantial costs for strategic stability and significantly increase the risk of unintended escalation. This demonstrates that doctrinal asymmetry is not merely a theoretical disagreement but an operational variable with tangible consequences during real crises.

Increased Probability of Miscalculation and Algorithmic Misperception

One of the most dangerous consequences of speed–safety asymmetry and compressed decision-making time is the substantial increase in the probability of miscalculation and algorithmic misperception during crises. When AI-enabled systems operate at speeds beyond the capacity for human evaluation and the two sides maintain

different levels of human control, the possibility of misinterpreting the adversary's intentions, actions, or even technical failures increases considerably. This danger becomes particularly concerning at flashpoints such as the Taiwan Strait and the South China Sea and in the context of the entanglement of conventional and nuclear systems. (6, 7, 19)

Algorithmic misperception refers to a situation in which algorithms used for data processing, target identification, or early warning incorrectly interpret an adversary's action or signal as hostile, preemptive, or indicative of an imminent attack. The speed of machine reaction severely constrains opportunities for humans to correct such misinterpretations and can trigger a chain of escalating actions. (5, 6)

Principal Flashpoints

The Taiwan Strait and the South China Sea are particularly prone to misperception because of their high concentrations of military forces, geographical proximity, the involvement of multiple actors, and their vital importance to both major powers. In these regions, conventional military movements, exercises, aerial interceptions, or even simple technical malfunctions could be interpreted by opposing early-warning algorithms and automated recognition systems as definitive evidence of an attack. (7, 12)

In the highly tense environment of the Taiwan Strait, the extraordinary speed with which machines can respond to sensor data may effectively close the temporal window required for careful assessment, political consultation, and activation of emergency communication channels. An algorithmic misperception could initiate actions whose consequences become difficult for humans to contain. Similar conditions exist in the South China Sea, where tactical encounters between vessels, freedom-of-navigation operations, and aerial interceptions can readily escalate into an uncontrollable crisis when combined with high-speed data processing. (6, 12)

The speed–safety asymmetry intensifies this danger. If one side's systems operate with higher autonomy and less human supervision, the opposing side, which remains bound by human-approval requirements, may interpret the adversary's faster actions as definitive evidence of aggressive intent and respond with preemptive measures. This dynamic lowers the crisis threshold and increases the likelihood that tactical tensions will escalate into strategic confrontation.

The Entanglement of Conventional and Nuclear Systems

The integration of, or shared dependence on, certain intelligence, command, warning, and detection infrastructures can create or intensify the risk of entanglement. The danger of miscalculation acquires strategic significance when conventional command, control, communications, and intelligence systems are entangled with nuclear systems. James Acton argues that cyber networks, early-warning satellites, and modern radar systems are increasingly shared between conventional and nuclear forces. (19) If, during a limited conventional conflict, a cyberattack or algorithmic disruption is directed against an adversary's radar or dual-use network, the targeted state may interpret the action not as a tactical measure but as preparation for a preemptive nuclear strike and an attempt to disable its second-strike capability. (6, 19)

Entanglement blurs the boundary between conventional and nuclear conflict and increases pressure toward escalation. Under conditions in which decision-making time is severely compressed and algorithms play a central role in processing warning signals, opportunities to correct misperceptions before crossing irreversible thresholds decline. The speed–safety asymmetry further magnifies this danger because the side operating more rapidly and

with less human control may inadvertently generate signals that the other side interprets within a worst-case framework. (5, 19)

Overall, the increased probability of miscalculation and algorithmic misperception constitutes one of the most serious consequences of speed–safety asymmetry for strategic stability. Flashpoints such as the Taiwan Strait and the South China Sea intensify this danger, while the entanglement of conventional and nuclear systems increases the possibility that tensions could spread from the tactical to the strategic level. In such an environment, preserving crisis stability increasingly depends on the ability to manage algorithmic ambiguity and maintain effective communication channels.

Intensification of the Risk of Instantaneous War

The speed–safety asymmetry, the compression of decision-making time, and the increasing probability of algorithmic misperception collectively intensify the risk of "instantaneous war." Instantaneous war refers to conditions in which hostile algorithmic systems or semi-autonomous platforms become trapped in an escalating cycle of action and reaction at speeds beyond the capacity for deliberate human intervention, transforming manageable tension into large-scale military confrontation within fractions of a minute or even seconds. (5) This danger is no longer merely a theoretical scenario but a logical consequence of combining machine speed with algorithmic ambiguity and reduced opportunities for human control.

In the context of U.S.–China competition—where one side places comparatively greater emphasis on speed and autonomy while the other emphasizes human control—conditions become more conducive to the emergence of such cycles. Two principal mechanisms intensify this danger: algorithmic escalation mechanisms and automation bias.

Mechanisms of Algorithmic Escalation

Algorithmic escalation occurs when autonomous or semi-autonomous systems on opposing sides react to one another's signals or actions without continuous human approval, causing these responses to expand in a chain-like fashion at the speed of data transmission. Drawing on the analogy of financial-market flash crashes, Paul Scharre demonstrates that the interaction of algorithms, each operating according to its own optimization logic, can generate systemic collapse even in the absence of any initial intention to produce catastrophe. (5)

On the battlefield, this mechanism may manifest in multiple forms. An aerial interception, naval maneuver, or even minor communications disruption may be interpreted by the adversary's early-warning algorithms as evidence of aggression. The machine's rapid response to this interpretation generates a new signal for the first side, thereby initiating an escalating cycle. Because the speed of such interaction may exceed the capacity for human processing and intervention, opportunities for de-escalation or correction of misperception may disappear. (6, 7)

The speed–safety asymmetry reinforces this mechanism. The side operating with greater autonomy and less human supervision can respond more rapidly, while the side bound by human approval finds itself at a temporal disadvantage. This asymmetry may create incentives for the faster side to establish battlefield realities more rapidly while simultaneously increasing pressure on the slower side to lower its control standards. (5, 12) The result is an increased likelihood of unintentionally crossing irreversible thresholds.

The Role of Automation Bias

Automation bias is one of the most significant human-organizational factors intensifying the risk of instantaneous war. It refers to the tendency of operators and commanders to overaccept the recommendations or decisions of automated systems and to become less willing to challenge them. (5) When humans observe successful autonomous-system performance over extended periods, they may gradually develop excessive confidence and, during critical moments characterized by high cognitive load and severe time pressure, lose either the capacity or motivation to intervene.

During a real crisis, this bias can have lethal consequences. Even in systems nominally characterized by a human-on-the-loop arrangement, meaningful human control may effectively disappear if operators refrain from vetoing machine decisions because of excessive trust or temporal pressure. (5, 7) The speed–safety asymmetry further magnifies this danger because the side facing temporal pressure from a faster adversary has greater incentives to rely on rapid machine-generated recommendations.

The combination of algorithmic escalation mechanisms and automation bias creates a dangerous pathway toward instantaneous war. Algorithms may initiate escalating cycles at machine speed, while humans affected by automation bias may be unable or unwilling to halt those cycles in time. Under such conditions, the boundary between human control and machine autonomy effectively collapses, rendering crisis stability highly fragile. (5, 6)

Overall, the intensification of the risk of instantaneous war constitutes one of the most serious consequences of the speed–safety asymmetry for strategic stability. Algorithmic escalation mechanisms enable the emergence of escalating cycles beyond effective human control, while automation bias weakens the capacity for meaningful human intervention even in systems that remain nominally supervised. In the context of U.S.–China competition, this combination significantly increases the risk that tactical tensions could transform into unintended strategic conflicts.

Weakening of Traditional Arms-Control Regimes

Traditional arms-control regimes, which are a legacy of the Cold War and played an important role in preserving strategic stability during previous decades, were built on foundations such as physical observability, the counting of warheads and delivery platforms, and mechanisms of reciprocal inspection and verification. Treaties such as START relied on the relative transparency generated through physical inspections and thereby contributed to greater predictability in state behavior. The emergence of artificial intelligence, with its distinctive characteristics—including its software-based nature, highly dual-use applications, dispersion across cloud environments and commercial servers, and rapid update cycles—has confronted this classical paradigm with conceptual and operational limitations. (9, 20) Consequently, the effectiveness of traditional regimes has declined, while the international normative vacuum surrounding military artificial intelligence has deepened.

The Software-Based and Dual-Use Nature of Artificial Intelligence

Unlike conventional or even nuclear weapons, artificial intelligence algorithms are software-based and intangible. They consist of lines of code that can be distributed anywhere in the world, trained on commercial servers or in civilian data centers, and updated rapidly. In practice, distinguishing precisely between civilian applications of such

software—such as supply-chain optimization or medical image processing—and lethal military functions, such as target recognition or drone guidance, is extremely difficult and at times impossible. (9, 20)

This dual-use nature fundamentally undermines traditional verification practices.

Remote sensing alone faces serious limitations in reliably identifying the ultimate purpose for which a data center is training a model, particularly in dual-use environments, and cannot confidently determine whether a large data center is training a model for commercial applications or optimizing military algorithms. Furthermore, traditional export controls, which focused on physical components, face serious limitations in regulating the transfer of algorithmic knowledge, pretrained models, and even open-source code. (12, 15)

Structural differences between the two countries further intensify this challenge. China's deliberate integration of commercial and defense sectors, on the one hand, (21) and the Pentagon's deep dependence on innovation from private technology companies, on the other, create significant political, legal, and operational barriers to targeted monitoring or restriction. (9) Under such conditions, traditional regimes based on counting and physical inspection are no longer effective instruments for managing algorithmic risks.

The International Normative Vacuum

An inevitable consequence of this situation is the substantial reduction in the feasibility of concluding restrictive treaties or even verifiable codes of conduct governing military artificial intelligence. Multilateral discussions within the framework of the United Nations Convention on Certain Conventional Weapons have clearly demonstrated the depth of disagreement and the relative deadlock that currently exists. A group of states and a substantial segment of civil society have called for preventive prohibitions or highly restrictive limitations on systems capable of selecting and engaging targets without meaningful human control. By contrast, major military powers, including the United States and China, emphasize the need to preserve military freedom of action, focus on interpreting and applying existing principles of international humanitarian law, and oppose preventive bans. (9, 20)

The absence of agreement even on precise definitions of concepts such as "autonomy," "meaningful human control," and acceptable thresholds of risk has produced a relative deadlock at the multilateral level. Intense bilateral competition between China and the United States has also reduced both sides' incentives to accept stringent normative constraints. Fear of falling behind in military capabilities has led both actors to prioritize freedom of action and development speed, further narrowing the space for the emergence of new restrictive norms. (9, 12)

This normative vacuum itself functions as an accelerator of crisis dynamics. Unless new concepts are developed for managing algorithmic risks and minimum levels of transparency, common red lines, and confidence-building mechanisms are established, unrestricted competition is likely to continue and strategic stability will become increasingly vulnerable.

Overall, the weakening of traditional arms-control regimes is one of the structural consequences of the emergence of military artificial intelligence. The software-based and dual-use nature of this technology has undermined the foundations of classical verification and deepened the international normative vacuum. In the absence of new regimes adapted to the characteristics of the algorithmic age, the risks of an accelerated arms race and unintended crisis escalation are likely to increase.

Discussion and Analysis

The analysis of military artificial intelligence competition between the United States and China demonstrates that the issue is not limited to the level of technological advancement or the capabilities of particular systems; rather, the manner in which the two sides address development speed, levels of autonomy, safety, and human control is equally significant. In this context, what may be termed the “speed–safety asymmetry” refers to a condition in which the speed of developing and employing AI capabilities may not correspond to the time and requirements necessary for testing, evaluation, control, and human intervention. The importance of this issue increases when military decisions are made during crises under severe time constraints for human assessment.

In the case of the United States, official Department of Defense policy emphasizes the need for appropriate levels of human judgment in the employment of autonomous systems. (9) Such an approach can enhance safety and facilitate human oversight, but under certain circumstances it may also make the development or deployment of particular capabilities more cautious and time-consuming. With regard to China, part of the existing literature emphasizes investment in artificial intelligence, autonomous systems, robotics, and swarm-based systems. (14) However, the available evidence alone does not justify the conclusion that China prioritizes speed over safety in every domain. Limitations in publicly available information regarding actual Chinese military programs and capabilities also require such comparisons to be undertaken cautiously.

The primary significance of this potential difference becomes clearer when examined from the perspective of temporal pressure. The faster AI systems are able to process data and recommend or execute military responses, the more limited the opportunities may become for human decision-makers to independently assess the situation. Under such conditions, the side that preserves more extensive control procedures may become concerned that slower decision-making during a crisis could result in the loss of operational initiative. Conversely, increasing speed in order to compensate for this disadvantage may itself increase the risk of errors and rushed decision-making. This issue is especially important in environments such as the Taiwan Strait, where crisis and military confrontation are plausible. (5, 6)

Another important consequence is the increased likelihood of algorithmic misperception and excessive reliance on AI-generated outputs. Intelligent systems also generate decisions or recommendations on the basis of data that may be incomplete, ambiguous, or inaccurate. Under normal circumstances, humans can review the system’s output and reject it where necessary; however, as decision time becomes shorter, opportunities for independent evaluation also decline. Consequently, operators may place excessive trust in system outputs and lose the time necessary to challenge them. (5) Therefore, the mere presence of a human in the decision-making process does not necessarily constitute “meaningful human control”; the quality and timing of human intervention are equally important.

Another issue concerns the possibility that risks may spill over from the conventional domain into the strategic and nuclear domains. This does not mean that conventional and nuclear systems are necessarily integrated into a single operational structure. Rather, the principal concern is that during a crisis, an action or error associated with conventional AI systems may be interpreted by the adversary as evidence of a broader or strategic threat. Under conditions of limited decision time and incomplete information, such an interpretation may facilitate crisis escalation. (6) From this perspective, artificial intelligence is significant not necessarily as a direct cause of war, but as a factor capable of shortening certain pathways toward unintended escalation.

From a theoretical perspective, this process can be explained through the security dilemma and game theory. An increase in the speed, autonomy, or processing capabilities of one side may be perceived by the other as threatening and induce it to develop comparable capabilities. Consequently, an action initially undertaken to increase security may generate greater mutual insecurity. In the present study, game theory is employed primarily as a conceptual framework for explaining this interaction rather than as a formal model intended to demonstrate a specific equilibrium or dominant strategy.

On this basis, a potential causal chain can be identified: increased speed and autonomy → reduced time for human decision-making → increased temporal pressure and ambiguity → increased probability of misperception and excessive reliance on algorithms → reduced opportunities for effective human intervention → increased risk of unintended crisis escalation. This chain demonstrates that the principal issue is not simply that technology is becoming "faster," but that speed becomes strategically consequential when technological processes outpace the capacity of humans and decision-making institutions to assess, control, and manage crises.

Within this framework, "instantaneous war" is better understood not as an inevitable outcome but as an extreme scenario arising from the compression of crisis decision-making time—a situation in which system reaction speeds increase to such an extent that insufficient opportunity remains for human intervention and assessment. Therefore, the findings of this study do not demonstrate that artificial intelligence necessarily increases the probability of war. Rather, they indicate that certain characteristics of the technology can generate conditions in which errors, misperceptions, and unintended escalation occur more rapidly.

From this perspective, reducing these risks does not necessarily require halting the development of military artificial intelligence. More importantly, mechanisms are needed to align technological speed with crisis-management capacity. Establishing dedicated communication channels for incidents and errors involving algorithmic systems, preserving meaningful human control in sensitive domains, defining minimum limits on the use of artificial intelligence in strategic systems, and employing selective transparency concerning safety procedures and human-control arrangements could constitute elements of such mechanisms.

Overall, the central issue in U.S.–China military AI competition cannot be reduced simply to an opposition between "speed" and "safety." The more important question is whether human and institutional crisis-management mechanisms can remain aligned with the speed at which technology is developing and being employed. If system decision-making and reaction speeds exceed human capacities for assessment and intervention, even a limited error may acquire broader consequences during a crisis. Accordingly, preserving strategic stability depends less on halting technological competition than on establishing mechanisms capable of controlling and managing the risks generated by that competition.

Conclusion

The expansion of military artificial intelligence is not limited to the development of more advanced weapons and systems; it is also transforming the manner in which decisions are made during crises. The examination of U.S.–China competition in this study demonstrates that the key issue is not merely the level of technological advancement, but also how an appropriate balance can be established between the speed of AI development and use, on the one hand, and safety, human control, and risk management, on the other.

One of the major findings of this study concerns the growing importance of time in military decision-making. The more rapidly AI systems are able to process information and recommend or execute appropriate responses, the

less time remains for humans to review information and make independent decisions. Under such circumstances, the mere presence of humans in the decision-making process is insufficient; what matters is whether humans genuinely have sufficient time to review a system's decision and prevent its execution when necessary. Human control is therefore meaningful only when it remains compatible with the operational speed of the systems involved.

Another issue that cannot be overlooked is the possibility of error and misperception during crises. Artificial intelligence systems depend on data that may be incomplete, ambiguous, or inaccurate. If outputs derived from such data are produced in circumstances where opportunities for review and correction are extremely limited, the likelihood of excessive reliance on algorithmic decisions or recommendations increases. In a crisis between major powers, such an error may not remain merely a technical malfunction and, if misinterpreted by the opposing side, could contribute to escalation.

The findings of this study indicate that an asymmetry between technological speed and safety requirements can constitute one factor contributing to reduced strategic stability. This conclusion does not imply that the development of artificial intelligence will directly or inevitably lead to war. What is important is the pathway that begins with increasing system speed and autonomy and, in the absence of appropriate control and communication, may lead to reduced decision-making time, increased risk of error, and ultimately unintended crisis escalation. Within this framework, "instantaneous war" should be understood as an extreme scenario—a condition in which system reaction speeds increase to such an extent that insufficient time remains for human intervention and assessment.

Accordingly, solutions cannot be sought merely in halting the development of military artificial intelligence. The more important requirement is that mechanisms for controlling and managing the associated risks develop in parallel with technological advancement. Establishing communication channels for incidents involving algorithmic systems, preserving meaningful human control in sensitive domains, defining minimum standards for the use of artificial intelligence in strategic systems, and increasing transparency concerning safety procedures may contribute to reducing these risks.

Overall, the principal challenge in U.S.–China military AI competition lies in the gap that may emerge between the speed of technological performance and the capacity for human decision-making. The wider this gap becomes, the more difficult crisis management will be. Therefore, preserving strategic stability in the age of artificial intelligence depends not only on possessing advanced technologies but also on the extent to which states can establish rules, constraints, and communication mechanisms commensurate with the speed and complexity of these technologies.

Acknowledgments

We would like to express our appreciation and gratitude to all those who helped us carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

All ethical principles were adhered in conducting and writing this article.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

1. Von Neumann J, Morgenstern O. Theory of Games and Economic Behavior: Princeton University Press; 1944.
2. Nash J. Equilibrium Points in N-Person Games. Proceedings of the National Academy of Sciences. 1950;36(1):48-9.
3. Jervis R. Cooperation under the Security Dilemma. World Politics. 1978;30(2):167-214.
4. Cave S, ÓhÉigeartaigh SS. An AI Race for Strategic Advantage: Rhetoric and Risks. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society2018. p. 36-40.
5. Scharre P. Army of None: Autonomous Weapons and the Future of War: W. W. Norton & Company; 2018.
6. Horowitz MC. When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability. Journal of Strategic Studies. 2019;42(6):764-88.
7. Payne K. I, Warbot: The Dawn of Artificially Intelligent Conflict: Oxford University Press; 2021.
8. Geist E, Lohn AJ. How Might Artificial Intelligence Affect the Risk of Nuclear War? Santa Monica, CA: 2018.
9. Scharre P. Four Battlegrounds: Power in the Age of Artificial Intelligence: W. W. Norton & Company; 2023.
10. Schelling TC, Halperin MH. Strategy and Arms Control: Twentieth Century Fund; 1961.
11. Allen JR, Husain A. On Hyperwar. Proceedings of the US Naval Institute. 2017;143(7).
12. Stokes J, Sullivan A, Greene N. U.S.-China Competition and Military AI. 2023.
13. Burdette Z, Wilson B, Frelinger DR, Demelash H, Arnold J. Exploring AI's Operational Implications in a Future U.S.-China Conflict: Initial Observations from the Camp(ai)gn Wargame. 2026.
14. Kania EB. Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power. 2017.
15. Boulanin V, Verbruggen M. Mapping the Development of Autonomy in Weapon Systems. 2017.
16. Engstrom J. Systems Confrontation and System Destruction Warfare: How the Chinese People's Liberation Army Seeks to Wage Modern Warfare. 2018.
17. Kania EB. Artificial Intelligence in China's Revolution in Military Affairs. Journal of Strategic Studies. 2019;44(4):515-42.
18. Work RO, Grant G. Beating the Americans at Their Own Game: An Offset Strategy with Chinese Characteristics. 2019.
19. Acton JM. Escalation through Entanglement: How the Vulnerability of Command-and-Control Systems Raises the Risks of an Inadvertent Nuclear War. International Security. 2018;43(1):56-99.
20. Maas MM. How Viable Is International Arms Control for Military Artificial Intelligence? Defense & Security Analysis. 2019;35(2):126-41.
21. Kania EB. "AI Weapons" in China's Military Innovation. 2020.