

Reckoning with International Relations and Strategic Security in the Age of AGI

Intelligence Competition, Deterrence, Crisis Stability, and the Boundary of Shared Survival

Reckoning with International Relations and Strategic Security in the Age of AGI

Intelligence Competition, Deterrence, Crisis Stability, and the Boundary of Shared Survival

International Relations / Strategic Security Independent Professional Research Paper V1.2.1

Author: Zijun Fu | Produced by: Civilization Leap Research Group

Evidence cutoff: August 26, 2026

Abstract

What advanced artificial intelligence entering international security truly changes is not merely weapon performance, but how states perceive threats, form judgments, transmit signals, allocate preauthorization, organize alliances, and decide on action. International relations and strategic security therefore cannot be replaced by the model-level question of whether AI is safe. A technically more reliable system does not automatically mean more stable relations among competing states. Under particular conditions, one side's shortening of its decision chain to gain an advantage may induce the other to accelerate, increase preauthorization, or shorten windows for verification, communication, and de-escalation; it may also prompt backing down, independent verification, oversight, refusal to use AI, or shared guardrails. Conversely, better sensing, simulation, early warning, defense, and crisis communication may reduce misperception and expand the time available for correction. AI's strategic consequences are not a one-way technological destiny. They depend on the combined effects of capability characteristics, organizational adoption, mutual expectations, alliance structures, strategic postures, legal institutions, and adversary responses.

Drawing on established traditions in international relations and strategic studies concerning the security dilemma, deterrence, crisis stability, alliances, signaling, misperception, arms races, and organizational processes, this article distinguishes four argumentative statuses. Category A comprises relatively stable historical, institutional, and actual deployment facts. Category B comprises unsettled competing explanations of whether AI strengthens or weakens deterrence stability, increases transparency or uncertainty, augments humans or drives automated escalation. Category C comprises cross-domain structural judgments such as "AI capability → strategic deployment → organizational processes → decision time → adversary expectations → behavioral adjustment → crisis stability → shared-survival risk." Category F comprises the author's publicly assumed normative commitments. The article does not turn category A facts directly into the inevitability of war, nor does it treat shared survival as an automatic justification for world sovereignty.

Its central judgment is this: what international security in the AGI era most needs to prevent is not a particular state's possession of stronger intelligence in itself. It is the gradual transformation of "faster, more automated, and unable to afford falling behind" into a shared default under incomplete information, mutual suspicion, and competitive pressure, such that capabilities intended to increase security jointly compress the space for correcting errors under particular conditions. This outcome must be empirically tested; it cannot be inferred automatically from unilateral performance or adoption. Strategic stability must therefore refocus from unilateral performance to the conditions of interaction. "Human in the loop" must also be tested further: do humans actually have sufficient information, time, authority, and organizational support to change outcomes? Here, shared survival is only a category F floor that no strategic competition may breach. It does not cancel legitimate national security interests, require unilateral dismantling of defenses, presuppose world government, or issue professional judgments on behalf of international law, ethics, political science, or computer science.

Keywords: AGI era; international relations; strategic security; military applications of artificial intelligence; deterrence; crisis stability; security dilemma; strategic autonomous systems; decision time; shared survival

Abstract

The strategic significance of advanced artificial intelligence does not lie only in improved weapon performance. AI is entering the processes by which states perceive threats, fuse intelligence, generate options, allocate authority, signal resolve, coordinate alliances, and act under time pressure. A technically reliable system does not therefore imply a strategically stable interaction. Under some conditions, one actor's attempt to shorten its decision cycle may induce rivals to accelerate, expand preauthorization, or shorten verification, communication, and de-escalation windows; under others, it may prompt restraint, independent verification, oversight, refusal, or reciprocal guardrails. Better sensing, simulation, warning, defensive coordination, and crisis communication may also reduce misperception. The effect of AI on strategic stability is therefore conditional rather than technologically predetermined.

This article builds from the established literatures on the security dilemma, deterrence, crisis bargaining, alliances, signaling, misperception, arms competition, command and control, and organizational failure. It distinguishes established facts and formal institutional materials (A), competing and unsettled explanations (B), cross-domain structural hypotheses requiring empirical validation (C), and the author’s explicit normative commitments (F). Its central claim is that the key unit of analysis is strategic interaction, not the AI system alone. Shared survival is treated only as a normative floor: it does not erase legitimate state security interests, authorize unilateral disarmament, create world sovereignty, or substitute for legal and ethical judgment.

Keywords: AGI era; international relations; strategic security; military AI; deterrence; crisis stability; security dilemma; autonomous systems; decision time; shared survival

Research Positioning, Scope, and Evidentiary Discipline

This is a professional research paper in international relations and strategic security. It examines how states, alliances, and military organizations alter one another's behavior under competition, incomplete information, and high-consequence conditions as AI capabilities enter intelligence, early warning, decision support, cyber operations, autonomous systems, command and control, and infrastructure. Its primary object is interaction, not an isolated model. Whether a model is reliable and how system assurance should be achieved must be reassessed by artificial intelligence and computer science. The legality of force and the application of international humanitarian law must be reassessed by law and international law. The legitimacy of public power and authorization of sovereignty must be reassessed by political science. Which harms can be justified must be reassessed by ethics. Within its own evidentiary scope, strategic studies answers only this: once these capabilities enter competitive interaction, how will they change deterrence, crisis stability, escalation control, and common security?

Throughout, the article uses four argumentative statuses: A, B, C, and F. Category A includes only relatively stable historical facts, established research, formal policies, and verifiable deployment status. Category B includes competing theories, contested mechanisms, and empirical results whose external validity remains unsettled. Category C includes cross-domain causal connections, future scenarios, and institutional stress tests. Category F comprises the author's publicly assumed normative commitments. A/B/C/F are not evidence grades. They remind readers what argumentative responsibility a statement bears. A cannot automatically generate B, C, or F; nor may F masquerade as a fact of strategic science.

Status	Typical content in this article	What it can support	What it cannot automatically establish
A Relatively stable facts / formal materials	Classical security studies, historical crisis materials, formal military AI policies, verified scope of application	Factual judgments concerning specified objects, times, and conditions	That AI necessarily causes war, or that a particular institution is necessarily legitimate
B Competing explanations	AI strengthens / undermines stability; transparency / new uncertainty; human-machine augmentation / automated escalation	Comparable mechanisms and empirical tests	A universal conclusion on which the strategic studies community has converged
C Cross-domain structural judgments	Speed competition → preauthorization → possible compression of verification windows; shared models → correlated risks within alliances	Falsifiable strategic hypotheses and stress tests	Historical facts that have already occurred

Status	Typical content in this article	What it can support	What it cannot automatically establish
F Normative commitments	Shared survival must not be compromised; highly irreversible decisions must retain conditions for responsibility and correction	An explicit account of why the article is concerned with certain risks and directions	A substitute for international legal validity, political authorization, or ethical judgment

Table 1. A/B/C/F argumentative statuses and the boundaries of strategic studies

The Minimal Interface with Civilization Leap Theory

Readers unfamiliar with Civilization Leap theory need not accept it in advance. This article uses only a minimal interface: reality unfolds through causality, not through disciplines; disciplines pursue truth separately, while civilization bears consequences collectively. International relations and strategic security can explain why states compete, how deterrence works, and why crises escalate or are contained. They cannot acquire final authority over legal, ethical, technical, or political questions merely because the problem ultimately concerns civilizational survival. Nor can a civilizational perspective predetermine the answers of strategic studies in reverse.

This article's category F floor is that shared survival is a civilizational boundary no strategic competition may breach. This adds an indispensable dimension of consequences to strategic choices; it neither declares national security interests illegitimate nor derives unified world power directly from "humanity as a whole." States must still address real threats, alliance responsibilities, and defense needs. If a common-security arrangement imposes unacceptable unilateral vulnerability on one side, it too will be difficult to sustain strategically. Shared survival and national realities must both be made visible, rather than canceling each other out.

Contents

1. Introduction When Intelligence Enters Strategic Interaction, Security Is No Longer a Matter of Unilateral Capability	
2. The Established Professional Legacy What International Relations and Strategic Security Have Established	
3. Contemporary Discontinuities Which Inputs to Strategic Interaction AI Is Changing	
4. Intelligence Competition and the Security Dilemma Why Individually Rational Choices May Produce Greater Collective Insecurity	
5. Deterrence Why Faster, More Accurate, and More Powerful Does Not Automatically Mean More Stable	
6. Crisis Stability Time Compression, Verification Windows, and Escalation Control	
7. Signaling, Attribution, and Deception More Information Does Not Mean Less Uncertainty	
8. Autonomous Systems and Decision Support From System Autonomy to Strategic Interaction	
9. Alliances and Interoperability Shared Capabilities Also Generate Shared Dependencies	
10. AI Infrastructure and Strategic Dependence Compute, Chips, Cloud, Data, and Supply Chains	
11. Nuclear Risk and Strategic Stability The Ultimate Irreversibility Stress Test	
12. Capability Diffusion and Nonstate Actors Diffusion Does Not Mean Equalization	
13. From Military Advantage to Common Security Which AI Uses May Improve Stability	
14. What International Governance Is and Is Not Establishing	
15. The Boundary of Shared Survival Where the Civilizational Perspective Enters and Where It Stops	
16. What Evidence Could Change This Article The Strongest Objections, Defeat Conditions, and Research Agenda	
Conclusion When Greater Speed Becomes Possible, Slower Verification Becomes More Valuable	
Further Reading and Joint Research Entry Points	
Appendix A Register of Source–Claim Boundaries	
Appendix B Strategic Stress-Test Matrix	
Declaration of Generative AI Use	
References	
Version Note	

1. Introduction | When Intelligence Enters Strategic Interaction, Security Is No Longer a Matter of Unilateral Capability

Security has never been a property possessed by one actor alone. Baldwin reminds us that discussion of “security” must specify whose values are protected against which threats and at what cost. The security dilemma developed by Herz and Jervis further indicates that measures one side takes to improve its own security may inadvertently reduce another's security, inducing countervailing armament and suspicion.[1–3] This established legacy has not become obsolete in the AI era; it has become more important. AI's strategic effects are rarely determined by one side's performance alone. They depend on how adversaries interpret, anticipate, imitate, and accelerate.

In 2024, the United Nations General Assembly adopted A/RES/79/239, treating AI in the military domain and its implications for international peace and security as a dedicated issue. It explicitly noted that states were integrating AI into a wide range of military applications and emphasized the applicability of international law throughout the lifecycle of relevant AI capabilities.[27] The Secretary-General's 2025 report A/80/78 compiled submissions from member states, international organizations, the International Committee of the Red Cross, research institutions, and industry, presenting opportunities and challenges extending far beyond autonomous weapons. The Assembly's 2025 resolution A/RES/80/58 subsequently decided to convene three days of open informal exchanges in 2026.[28–30] Military AI is thus no longer merely a technological forecast; it has become a formal international security agenda item.

Yet “entering the international agenda” does not mean “the risks have been proven.” UNIDIR's research on military AI emphasizes that AI can participate in eighteen military task categories beyond weapons, encompassing information processing, planning, logistics, and support functions. Technical feasibility, actual adoption, organizational absorption, and strategic consequences must still be assessed separately.[25] This article therefore rejects two simplifications: portraying every military adoption of AI as inevitably escalatory, and dismissing strategic risks as exaggeration because no single decisive piece of evidence yet exists. The intermediate chain is the real object of inquiry.

1.1 The Article's Central Question

The article asks: as AI continues to improve information processing, sensing, prediction, decision support, cyber operations, and some autonomous execution capabilities, while states remain in competition under incomplete information and concern over relative advantage, how will the security dilemma, deterrence, crisis stability, escalation control, alliance interaction, and conditions for shared survival change? This question requires distinguishing at least five levels: technical capability, organizational adoption, strategic posture, adversary perceptions, and interaction outcomes. None can directly substitute for another.

1.2 Three Equivalences That Cannot Be Assumed

Observed change	Common automatic inference	The bridge still requiring demonstration
A faster system	A more secure state	Speed can become security only if verification, responsibility, adversary responses, and de-escalation windows remain viable
More information	Less misperception	Data quality, deception, model calibration, organizational cognition, and attribution determine whether information actually reduces uncertainty
More autonomous weapons	More credible deterrence	Autonomy may improve responsiveness, but may also reduce controllability and increase misinterpretation and fears of losing control

Table 2. Three missing bridges between technical performance and strategic outcomes

2. The Established Professional Legacy | What International Relations and Strategic Security Have Established

2.1 The Security Dilemma: Security as an Outcome of Interaction

The central point of the security-dilemma tradition is not that “states are necessarily hostile.” Rather, even when actors primarily seek self-preservation, defensive choices may be read as threats if intentions cannot be credibly communicated, offensive and defensive postures are difficult to distinguish, and confidence in the future is lacking.[2–3] What AI changes first is therefore not whether states are self-interested, but whether offensive and defensive capabilities become easier to distinguish, whether advantages shift faster, and whether signals connecting intentions and capabilities become harder to interpret.

2.2 War and Crisis: Information, Commitment, and Misperception Remain Central

Fearon's rationalist explanations for war center on private information and incentives to misrepresent it, alongside commitment problems. Schelling understands coercion, deterrence, and manipulation of risk through mutual expectations.[5,7] Jervis, Allison, and subsequent organizational research further show that states are not frictionless, unitary rational actors: perceptual biases, bureaucratic procedures, organizational routines, and limited information can alter crisis choices.[8–12] These legacies mean that AI does not simply “add computing power to rationality.” It may alter information structures, or encode old organizational biases in faster processes.

2.3 Strategic Stability Is Not a Single Indicator

Strategic stability requires distinguishing at least the following: whether peacetime conditions continually generate arms-racing pressure; whether crises create incentives for preemptive or preventive action; whether conflict readily escalates across domains and levels once it begins; whether deterrent threats are credible yet remain controllable; and whether command and control can maintain authorization and identification under pressure.[13–19] Research on emerging technologies repeatedly indicates that the same technology can have opposite effects across these dimensions, and may stabilize one stage while destabilizing another.[13–15] Accordingly, this article does not construct an “overall AI strategic stability score.”

Dimension	Core question	Principal variables after AI enters
Peacetime / arms-race stability	Does fear of falling behind continually drive earlier deployment and arms expansion?	Adoption speed, relative capability, diffusion, visibility, offense–defense distinguishability
Crisis stability	Does a crisis generate preemptive pressure that “failing to act now will make things worse later”?	Early warning, decision time, false alarms, adversary posture, communication windows
Deterrence stability	Are threats and commitments credible and comprehensible?	Capability, intention, controllability, signal consistency, resilience of retaliation
Wartime escalation stability	Can actions remain confined to controllable levels after conflict begins?	Cross-domain coupling, automatic responses, target identification, organizational routines
Command-and-control resilience	Can the chain of authority identify errors, take over, stop, and repair?	Data, communications, permissions, redundancy, human-machine relations

Table 3. Dimension-by-dimension review of strategic stability

3. Contemporary Discontinuities | Which Inputs to Strategic Interaction AI Is Changing

AI is not the first technology to enter warfare. Radar, satellites, nuclear weapons, precision guidance, and cyber technologies have all changed information, firepower, and organization. The current discontinuity is that the

same family of general-purpose AI capabilities can enter intelligence fusion, target identification, cyber defense, logistics, simulation, information operations, decision support, and autonomous systems simultaneously. Many capabilities originate in civilian markets and can be rapidly updated through software. Horowitz therefore treats AI as closer to an “enabling technology” than a single weapon; its effects depend heavily on organizational absorption and complementary institutions.[16]

UNIDIR's 2023 report expands the military AI perspective beyond “autonomous weapons” to a broad range of tasks. NATO's revised 2024 strategy likewise places generative AI, dual-use civilian solutions, data, compute, energy, human-machine accountability, and interoperability within one policy framework.[25,36] NATO's 2025 annual report also records its completed procurement of the Palantir Maven Smart System NATO and testing in exercises to enhance data analysis, decision processes, and strategic responsiveness.[39] These are category A adoption facts. They establish only that AI is entering military organizations, not that it has already improved or worsened strategic stability.

3.1 Six Input Conditions Requiring Renewed Research

Input condition	Possible capability change	Strategic question still to be asked
Speed	Information processing, option generation, and task execution may accelerate	Are verification, authorization, and communication shortened in parallel?
Visibility	Sensing and data fusion may improve situational awareness	Does the adversary interpret stronger sensing as defense, preparation for attack, or a first-move advantage?
Uncertainty	Models can quantify probabilities and process vast amounts of data	Do data contamination, out-of-distribution situations, deception, and model errors create new uncertainty?
Autonomy	Some tasks can be completed with less immediate human input	Are authorization boundaries clear, and can humans still take over when anomalies arise?
Diffusion	Software and civilian capabilities can spread rapidly across organizations	Do capabilities actually diffuse equally, or do compute, data, talent, and organizational capacity stratify them anew?
Dual use	Civilian infrastructure can become a condition for military capability	Do commercial supply chains, cloud services, chips, and models become sources of strategic dependence?

Table 4. Contemporary discontinuities in the inputs to strategic interaction in the AGI era

3.2 “AGI” Is Not a Prerequisite for the Article's Argument

The “AGI era” here does not claim that AGI, under some unified definition, has already been achieved. It means that advanced AI is continually entering high-consequence decision chains, beginning to reorganize strategic interaction around the prospect that increasingly powerful intelligence capabilities can be widely invoked. When genuine AGI will emerge, and whether it possesses subjective experience or continuity of subjecthood, are not premises of this article's strategic conclusions. Whether current systems can be trusted, how they should be tested, and whether they are conscious remain matters for computer science, epistemology, and research on mind, respectively.

4. Intelligence Competition and the Security Dilemma | Why Individually Rational Choices May Produce Greater Collective Insecurity

When AI is seen as the shared underlying capability for economic, intelligence, cyber, unmanned-system, and command advantages, competitors face a practical incentive: even if they cannot establish whether “more AI” actually increases security, they may fear that nonadoption will first bring a relative disadvantage. This mechanism is consistent with the traditional security dilemma, but software diffusion, civil-military integration, and the pace of capability updates give it a new temporal structure.[3,16]

4.1 Competition Is Not Simply About Who Acquires AGI First

Strategic competition can occur at multiple intermediate levels: faster intelligence fusion, wider sensor coverage, cheaper unmanned systems, greater cyber automation, shorter chains of command, larger compute reserves, and faster model updates. Research focused solely on an “AGI arrival date” will miss the incremental competition already taking place. Conversely, calling all AI investment an “arms race” is also an exaggeration. What must actually be observed is whether one side's investment systematically changes another's threat perceptions and deployment choices, and whether this draws them into mutually reinforcing adoption cycles.

4.2 The AI Label Cannot Determine Offensive or Defensive Character

Jervis's offense–defense distinction and Garfinkel and Dafoe's research on how the offense–defense balance changes with investment scale both suggest that a technology's offensive or defensive orientation must be assessed relative to specific tasks, geography, adversaries, and levels of investment.[3,15] AI can improve cyberattacks, reconnaissance, and target searches, but also air defense, anomaly detection, maintenance, and resilience. “Stronger AI” cannot justify presuming offensive dominance, much less deriving a first-move advantage directly from abstract compute scale.

4.3 Category B/C Conditional Propositions: When Speed Competition May Become Institutional Competition

Existing evidence supports only some local links in the chain, not proof that the complete outcome has occurred. First, an adversary's learning that AI or automation is involved in high-consequence decisions may change its judgments of threats, credibility, accidents, and responsibility (P3a, B). Existing experiments and cognitive research show that algorithmic provenance affects national security experts' or military personnel's willingness to act, attribution of responsibility, and trust, but not in a single direction.[43,45–46] Second, these perceptions may elicit retaliation, backing down, independent verification, efforts to catch up, countermeasures, oversight, refusal to use AI, or shared guardrails; the direction depends on task, organizational, and crisis conditions (P3b, B–C). Research on Chinese and US perceptions of military AI and on automated nuclear threats suggests that anxiety over relative advantage, imitation, and changes in credibility can all drive responses. It does not establish that adversaries necessarily accelerate symmetrically.[44,47]

Third, the parties may jointly compress the time reserved for verification, communication, and political re-entry only when one side's process changes are understood by the adversary as shortening detection, decision, or action times and trigger further automation, preauthorization, alerting, or shortened authorization procedures, and those responses in turn change the other side's processes (P3c, C). Fourth, this joint compression may increase risks of misperception, accidents, or unintended escalation only under conditions such as low confidence in second-strike capability, opaque doctrine, weak independent verification, and fragile crisis communication (P3d, C). Better defenses, independent review, hotlines, reversible authorization, and shared guardrails may instead expand the time available for correction.[17,48–50]

What truly matters here is not how quickly a particular algorithm runs, but how costly the overall strategic interaction makes being “one step slower.” The article does not, however, assume that AI necessarily accelerates crises or that adversaries adapt symmetrically, and it does not claim that this complete chain has been empirically observed.

Directly adjacent to this article, Yuan (2026) has already discussed action–reaction pressures, adoption incentives, cyber risks, deterrence, and crisis decision-making in US–China military AI competition.[51] This article

does not claim that interactive time compression has been empirically observed, nor that it is the first to raise those broad questions. Its specific contribution is to separate “perception–response–shared time budget–crisis risk” into four propositions, P3a–P3d, each of which can be defeated independently, and to operationalize the jointly reserved time for verification, communication, and political re-entry as (T_V, T_C, T_P). It thus shifts the question from “can one side decide faster?” to “does reciprocal, multiround adaptation shorten the time for correction before consequences become irreversible?”

5. Deterrence | Why Faster, More Accurate, and More Powerful Does Not Automatically Mean More Stable

Deterrence is not capability possession itself. It is making an adversary believe that a particular action will bring unacceptable consequences while keeping threats, commitments, and limitations sufficiently credible.[6–7] Huth and Russett’s historical research shows that deterrence success and failure depend on specific relationships, local power conditions, and commitment structures, and cannot be summarized by an aggregate military-power indicator.[6] AI’s entry into deterrence therefore requires distinguishing at least capability, visibility, credibility, controllability, and interpretation.

5.1 Three Pathways Through Which AI May Improve Deterrence

First, better situational awareness and early warning may improve the survivability of retaliatory capabilities, making it harder for adversaries to believe a surprise attack could succeed. Second, simulation and decision support may help leaders compare more crisis pathways and identify misperceptions and alternatives. Third, cyber defense, anomaly detection, and unmanned performance of high-risk tasks may reduce vulnerability, so that a state need not act early out of fear of abruptly losing its capabilities.[20–24] These are stabilizing possibilities deserving serious research, not exceptions that technological pessimism should ignore.

5.2 Three Pathways Through Which AI May Also Weaken Deterrence

If the same sensing and analytical capabilities are perceived as able to locate an adversary’s concealed forces, disable command networks, or improve first-strike effectiveness, they may intensify fears that a window is closing. If adversaries consider automated responses to lack political controllability, threats may appear both more credible and more dangerous. Opaque model outputs and data sources may also make it impossible to tell whether another state has genuinely acquired new capabilities or merely believes that it has.[17–23] Deterrence stability therefore depends on changes in both capabilities and perceptions.

AI-related change	Possible stabilizing effect	Possible destabilizing effect
Faster early warning	Earlier attack detection and more resilient retaliation	False alarms enter decisions faster, compressing human review
More accurate target identification	Fewer mistaken strikes and more efficient defense	Adversaries fear concealed forces are easier to discover, generating vulnerability
Automated response	Reduced risk of paralysis after communications disruption	Reduced controllability, inadvertent triggering, or adversary misinterpretation as preauthorization
Simulation of more options	Support for comparing consequences	Shared model blind spots, false precision, and automation dependence
AI cyber defense	Greater system resilience	Similar AI can also support attacks, deception, and reconnaissance

Table 5. Bidirectional possibilities of the same AI capability for deterrence

6. Crisis Stability | Time Compression, Verification Windows, and Escalation Control

Crisis stability concerns whether, under severe tension, parties acquire incentives to move first because they fear that waiting will worsen their position. Speed is the AI-related variable most deserving caution and most easily exaggerated. Speed itself is not a risk. The real questions are which processes accelerate, which fail to acquire better evidence in parallel, and whether political decision-making still has time to re-enter.

6.1 Shorter Decision Time Is Not Always Better

During disasters, air defense, or cyber anomalies, faster identification can expand correction windows. In mutual strategic competition, greater speed may instead make waiting for verification substantially costly. Horowitz identifies greater operational speed and reduced human control as two key features through which autonomous weapons may affect crisis and strategic stability.[17] Yet Talmadge's Cold War case research cautions that new technologies are usually not independent sufficient causes of escalation; more often they are variables intervening in existing political and military conditions.[14] These arguments are compatible: speed warrants study, but cannot be turned into a law of automatic escalation.

6.2 Meaningful Human Control Is First a Question of Time

Formal "human-in-the-loop" arrangements are insufficient. If a person has only seconds to confirm a problem definition, evidence selection, target ranking, and action recommendation already completed by a model, that person may still click the button without truly changing the outcome. Strategically meaningful human control requires at least access to information proportionate to the consequences; understanding of uncertainty and alternative explanations; authority to refuse, delay, de-escalate, or refer a decision for higher review; an organization that does not punish all caution as "missing the window"; and the ability, at critical moments, to turn machine advice back into political judgment. DoD 3000.09 and several international political commitments emphasize appropriate human judgment, testing, and failure handling. Whether these institutional requirements work under real high-pressure conditions remains an empirical question.[37–38]

6.3 Decision Windows Should Be Treated as Strategic Resources

Traditional strategic studies emphasizes weapons, survivability, and communication chains. The AI era also requires studying "time available for verification, challenge, and communication" as an independent object. The shortest response time is not inherently advantageous. If faster responses also remove redundancy in judgment, a system may perform well under normal conditions yet become more fragile under anomalous ones. Conversely, deliberately retaining pauses, dual confirmation, cross-sensor checks, and crisis communication may sacrifice local speed while increasing overall stability. This trade-off requires wargaming, organizational experiments, and cross-national crisis research; the article cannot settle it on paper.

This article represents the shared time budget as a three-dimensional vector (T_V , T_C , T_P). T_V is the time still available, from the emergence of a critical anomaly or competing interpretation, to complete independent verification, generate counterexamples, and review authorization. T_C is the time still available for parties to check intentions, correct attribution, communicate limitations, and initiate crisis communication before action becomes irreversible. T_P is the time still available for political decision-makers authorized to bear the consequences to re-enter and change a trajectory already set by machines or organizations. The three dimensions must not be combined into an "overall strategic correction score," nor may an increase in one be traded for the failure of another.

The minimum criterion for mutual compression requires observation of bilateral or multilateral process feedback over at least two rounds: one party's process changes trigger an adversary response, which in turn changes the former's processes. At the same time, each party must experience a net reduction, relative to its own baseline, in at least one dimension common to all parties, and that dimension must not be restored by additional independent verification, effective communication, or political review. Unilateral acceleration, a shorter single round, reduced purely technical latency, or a shorter overall process does not suffice to demonstrate joint compression. Backing down, independent verification, oversight, refusal to use AI, and shared guardrails must be coded simultaneously as competing outcomes.

7. Signaling, Attribution, and Deception | More Information Does Not Mean Less Uncertainty

The difficulty in international crises has never been merely a lack of information. It also concerns who sends it, whether the other side believes it, whether incentives for deception exist, and whether actions are misattributed. Fearon's work on private information and incentives to misrepresent it, and Jervis's research on perception, both show that more information does not automatically eliminate strategic uncertainty.[5,8] As AI expands capacities to generate, filter, synthesize, and analyze information, it may improve visibility while creating new problems of trust.

7.1 Deepfakes Are Not the Only Problem

Synthetic audio and video may certainly affect crisis signals, but the deeper issue is the authenticity of the entire information chain: whether sensors have been deceived, data contaminated, models subjected to prompt injection or adversarial manipulation, provenance chains preserved, and decision-makers informed that a model is inferring rather than observing. NATO's revised AI strategy already identifies AI-enabled information operations and disinformation as concerns, emphasizing testing, data, and human-machine accountability.[36] Such policy concerns establish only that risks have been formally recognized, not that every manipulation works in real crises.

7.2 Improved Attribution May Also Change the Structure of Escalation

Cyber-conflict research has long challenged the proposition that anonymity necessarily destroys deterrence. Gartzke and Lindsay both argue that the strategic effects of cyber operations depend on targets, context, intelligence, and connections to real-world military action, rather than the "cyber" label itself.[40–41] AI may improve anomaly detection and attribution speed, but may also increase attack automation and the sophistication of deception. Strategic studies must compare net outcomes: in which settings does AI make attribution more reliable, and in which does it merely accelerate mistaken attribution?

7.3 Signals Must Remain Comprehensible to the Other Side

Both deterrence and de-escalation depend on the other side understanding what actions mean. If many decisions are generated by invisible models and internal organizational rules, an adversary may observe an action without knowing whether it represents political intent, an automatic response, a false alarm, or local command behavior. AI governance must therefore examine strategic interpretability as well as internal reliability: which capabilities, limitations, authorization structures, and crisis-communication information can be understood externally without disclosing critical secrets?

8. Autonomous Systems and Decision Support | From System Autonomy to Strategic Interaction

Autonomous weapons are an important subset of military AI, but do not encompass all of it. The ICRC describes autonomous weapons as systems that, after activation, can select targets and apply force without further human intervention, and continues to advocate new legal constraints on unpredictable systems and autonomous attacks against persons.[34] This is a legal and humanitarian position attributable to a specific institution. The article cannot present it as a unified legal conclusion already reached by the international community.

8.1 Autonomy Is Not an On-Off Switch

A more useful approach in strategic analysis is to disaggregate the action chain: who defines goals, selects data, establishes identification rules, permits the system to operate within particular geographic and temporal bounds, can modify the mission, confirms an attack, can stop it, and assumes responsibility for remediation after errors? A system can be highly autonomous in execution yet strictly controlled at the level of purposes and authority. Conversely, it can formally retain human confirmation while algorithms effectively predetermine almost all visible options.

8.2 Human Responsibility Still Requires Organizational Structures

International policy documents increasingly emphasize human responsibility. The US political declaration requires military AI to operate within a responsible human chain of command and control, and calls for legal reviews, defined uses, lifecycle testing, and safeguards against failures. REAIM 2026's Pathways to Action similarly identifies human judgment, risk assessment, testing and validation, and reducing misunderstanding, miscalculation, and unintended escalation as practical directions.[31,38] The strategic question for these principles is whether responsibility remains real amid speed, layers of outsourcing, and supply-chain dependence.

8.3 An Adversary Still Stands Between Tactical Reliability and Strategic Stability

An autonomous system may be highly reliable within a bounded task yet still change strategic interaction because of its scale, speed, deployment location, or what the other side imagines. Conversely, a technically imperfect system need not create the same pressure if its purpose is transparent, scope limited, errors controllable, and adversaries do not view it as a first-move instrument. Strategic studies cannot translate “technical reliability” directly into “international stability,” nor “technical flaws” directly into “inevitable escalation.”

Level of the action chain	Question to answer first	Disciplinary handoff
Purpose / mission definition	Who decides what military purpose the system should achieve?	Political / military authorization; this article analyzes only the consequences for strategic interaction
Data / identification	Which observations enter the system, and how are classifications formed?	AI / computer science and epistemology; strategic studies examines consequences for misinterpretation and signaling
Authority / scope	Within which times, areas, and target categories may it act?	Law / institutional authorization; strategic studies examines adversary expectations
Execution / autonomy	When may the system act on its own?	AI engineering; strategic studies examines speed, controllability, and escalation
Stopping / remediation	Who can halt, isolate, recover, and explain accidents?	Systems engineering / organization; strategic studies examines crisis de-escalation capability

Table 6. Action-chain components that must be distinguished before autonomous systems enter strategic interaction

9. Alliances and Interoperability | Shared Capabilities Also Generate Shared Dependencies

Alliance research has long emphasized that alliances provide security while generating twin fears of abandonment and entrapment. Internal commitments, capability differences, and strategic autonomy are not peripheral questions.[4] In the AI era, interoperability extends further into models, data, cloud services, interfaces, and software updates. NATO's revised AI strategy explicitly identifies alliance-wide AI system interoperability as an objective while emphasizing testing, standards, and responsible use.[36]

9.1 The Stability Benefits of Interoperability

Shared data standards, situational pictures, and decision support may improve alliance coordination, reduce misperception among friendly forces, reduce information silos, and improve collective-defense efficiency. These benefits have practical value for multinational air defense, maritime awareness, cyber defense, and logistics. Alliances therefore have ample reason to pursue AI interoperability.

9.2 Interoperability May Also Create Correlated Failures

If several allies depend on the same models, cloud services, data pipelines, or suppliers, local vulnerabilities, model biases, and policy changes may spread simultaneously. More complex still, capability asymmetries can alter responsibility within alliances: technically stronger members may in practice supply more of the basis for judgment, while weaker members still bear the political consequences of action. In such circumstances, “shared use” does not mean “shared understanding,” much less “shared authorization.”

9.3 Alliance AI Requires Two Levels of Visibility

The first is internal: allies know system capabilities, limitations, versions, and permissions. The second is external: adversaries can understand, to the necessary extent, what particular deployments mean, avoiding misinterpretation of routine interoperability upgrades as the sudden creation of offensive preauthorization. Transparency certainly has security limits, but complete unintelligibility also carries strategic costs. National security, arms-control, and legal specialists must determine the specific scope of disclosure.

10. AI Infrastructure and Strategic Dependence | Compute, Chips, Cloud, Data, and Supply Chains

Another feature of AI as an enabling technology is that military capabilities increasingly depend on civilian and commercial infrastructure: chips, data centers, cloud platforms, networks, model providers, software ecosystems, and energy. NATO's 2024 strategy explicitly includes civilian dual-use solutions, compute and energy requirements, and industrial and academic ecosystems among the conditions for defense AI.[36] “Strategic autonomy” thus concerns more than weapons stocks; it also concerns whether critical digital infrastructure can continue operating, be replaced, and recover.

10.1 Infrastructure Control Does Not Automatically Confer Sovereignty

Control of important cloud, model, or chip supply chains can create substantial de facto influence, but does not automatically confer legitimate public authority. That remains for political science and law to assess. Strategic security is concerned here with how dependence changes crisis vulnerability: could supply disruption deprive a state of critical capabilities during a crisis? Could commercial terms of service, cross-border data, and software updates become points of strategic friction?

10.2 Attacks on Infrastructure May Produce Cross-Domain Escalation

AI infrastructure often serves civilian economies and military organizations simultaneously. Actions against cloud services, data centers, communications, or chip supply chains may closely couple military value and civilian consequences. An adversary may also be unable to tell whether a cyber or physical action is a limited military measure, a broad economic attack, or preparation for greater escalation. International relations, cybersecurity, international law, infrastructure resilience, and ecological and energy expertise must therefore re-engage; “strategic necessity” cannot settle all these questions at once.

11. Nuclear Risk and Strategic Stability | The Ultimate Irreversibility Stress Test

The nuclear domain is this article's strictest stress test because the consequences of error may be irreparable. SIPRI's series of studies since 2020 repeatedly analyzes the bidirectional possibilities of AI entering early warning, ISR, command and control, unmanned platforms, and related conventional systems. Better sensing, maintenance, and decision support may strengthen resilience; premature adoption, false alarms, cyberattacks, capability misperception, and arms racing may also increase risk.[22–24] These studies largely draw on technical-strategic synthesis, regional expert workshops, and public sources. They are appropriate category B and policy-analysis materials, not a basis for a universal category A claim that “AI has already entered nuclear launch decisions.”

11.1 The Evidentiary Boundary in 2026

Saalmann's 2026 systematic survey of public sources on China, India, and the United States finds substantial differences in their openness concerning AI and nuclear and nuclear-related systems, and positions the work as a

baseline for future engagement and confidence-building measures.[24] This demonstrates precisely why current research requires restraint. Public materials can establish some AI-related functions, policies, and trends, but are generally insufficient to reconstruct the full internal algorithms, authorization arrangements, and actual nuclear decision processes of national NC3 systems. This article therefore does not claim to know of any state delegating final authorization of nuclear use to AI.

Equally important, on December 1, 2025, the UN General Assembly adopted A/RES/80/23, specifically addressing possible risks of integrating AI into nuclear weapons command, control, and communications systems. Pending the complete elimination of nuclear weapons, the resolution calls for maintaining human control and oversight of relevant systems incorporating AI technologies. It passed by 118 votes to 9, with 44 abstentions.[42] It establishes that AI and nuclear command, control, and communications have entered a distinct multilateral nuclear-risk agenda. It does not establish uniform practice among all nuclear-weapon states, nor create new treaty obligations.

11.2 The Nuclear Domain Most Needs Full Political Judgment, Not Merely a Human Signature

In a nuclear crisis, the real danger may lie less in a system launching directly than in its deciding ever earlier what is seen, which signals are considered credible, which options rank first, and how quickly a response is required. If political leaders choose only at the final step among a handful of options already compressed by machines, formal human authorization remains while the substantive space for judgment may narrow. An abstract slogan of “human control” cannot solve this. It requires command-and-control research, organizational simulation, crisis exercises, and audits of actual procedures.

Application stage	Potential benefit	Principal risk	Priority for validation
Early warning / situational awareness	Faster detection and fusion	False alarms, adversarial deception, false precision	Cross-checking different sources and retaining uncertainty
NC3 assistance	More efficient information organization and communications	Common-component failures, cyber infiltration, automation dependence	Redundancy, isolation, tiered authorization
Conventional counterforce capabilities	Improved localization and mission planning	Adversary fears of nuclear-force vulnerability and increased preemptive pressure	Strategic dialogue, limiting misinterpretation, transparency / confidence-building measures
Unmanned / autonomous platforms	Greater survivability and persistence in dangerous environments	Unclear deployment intentions, loss of control, or mistaken escalation	Limited mission scope, stopping capability, accident communication
Model-based simulation	Expanded comparison of options	Training-data bias and mistaking models for reality	Multimodel / human adversarial review and explicit failure conditions

Table 7. Bidirectional stress tests of AI and nuclear risk

12. Capability Diffusion and Nonstate Actors | Diffusion Does Not Mean Equalization

The replicability of AI software readily encourages the belief that capabilities will diffuse rapidly and all actors will soon approach the same level. Horowitz's framework for military technology diffusion and Goldfarb and Lindsay's analysis of the complementarity of data and judgment remind us that real military capabilities still depend on organization, data, talent, hardware, training, mission understanding, and integration capacity.[16,19] Open models or falling costs may therefore lower some barriers without necessarily eliminating differences among states and organizations.

12.1 Risks from Nonstate Actors Must Be Disaggregated by Task

AI may lower barriers to information gathering, code generation, linguistic manipulation, unmanned-platform planning, or target analysis, but the practical barriers vary enormously across tasks. Equating “access to a model” directly with “acquisition of strategic-level capability” generates alarmism. Conversely, ignoring certain low-cost, high-impact tasks because nonstate actors lack state-level infrastructure misses real risks. Research should assess specific tasks, resources, data, and action chains rather than settle the issue with the phrase “AI democratization.”

12.2 Diffusion Also Reshapes States' Crisis Decisions

When states cannot rapidly establish who launched an attack, or fear that nonstate actions will be mistaken for state behavior, attribution becomes part of escalation control. AI-generated content, proxy networks, and automated attack tools may further blur chains of responsibility. Re-engagement with cybersecurity, intelligence, law, and research on social information is essential here.

13. From Military Advantage to Common Security | Which AI Uses May Improve Stability

If this article merely listed risks, it would make the same kind of error as technological optimism, in reverse. What strategic studies must actually compare is which AI uses can increase verifiable defense, early warning, communication, and correction capabilities without creating greater mutual uncertainty. UNODA, REAIM, and many national policies discuss opportunities for AI to improve information processing, personnel safety, situational awareness, logistics, and decision support alongside the risks.[26,28,31]

13.1 Defensive, Distinguishable, and Verifiable Capabilities Are More Likely to Reduce the Dilemma

Security-dilemma research offers a direction: conditions for cooperation are generally better when defense has an advantage over offense and defensive postures are easier to distinguish from offensive ones.[3] AI's common-security value may therefore lie not in “everyone using less AI,” but in prioritizing and making visible applications that improve infrastructure resilience, accident detection, communication reliability, disaster response, and limited defense without significantly increasing first-strike capabilities. Which uses meet these conditions must be verified individually.

13.2 AI Can Preserve More Options, Not Merely Enable Faster Choice

One of the most valuable functions of high-quality decision support may be actively generating counterexamples, displaying uncertainty, comparing de-escalation pathways, flagging information gaps, and preserving delay options, rather than presenting leaders with a single “optimal action.” Goldfarb and Lindsay emphasize that cheaper prediction raises the relative importance of judgment.[19] Strategic systems may misuse AI if they optimize response speed alone. If used to expand leaders' understanding of uncertainty and alternatives, AI may enlarge rather than compress political judgment.

13.3 Arms Control and Confidence-Building Measures Can Also Use AI Tools

AI can help monitor public information, organize treaty data, identify anomalies, simulate verification arrangements, and support information processing for crisis hotlines. REAIM 2026 also includes voluntary confidence-building measures, sharing policy and risk-assessment information, and reducing misunderstanding, miscalculation, and unintended escalation among its pathways to action.[31] These tools do not automatically generate trust. Yet if they improve verifiability without concentrating sensitive judgments in black boxes that cannot be challenged, they may become stabilizing infrastructure.

14. What International Governance Is and Is Not Establishing

As of August 26, 2026, military AI governance has clearly moved from scattered initiatives toward multiple parallel layers. There is, however, no single globally unified, comprehensively legally binding regime covering every military AI issue. Describing this accurately matters: one cannot say that “the international community has no

rules,” nor treat political declarations, UN General Assembly resolutions, national policies, and treaty law as having identical legal force.

14.1 The UN General Assembly: The Issue Has Formally Become Global

A/RES/79/239 incorporated AI in the military domain and its implications for international peace and security as a dedicated issue in a General Assembly resolution, emphasizing the applicability of international law throughout the lifecycle of relevant AI capabilities and requesting a Secretary-General's report.[27] A/80/78 was submitted in 2025, bringing together diverse views. A/RES/80/58 subsequently decided to continue informal exchanges under UN auspices in 2026.[28–30] The Geneva meeting of June 15–17, 2026 has taken place, covering terminology, capability trends, the lifecycle, technological convergence, international cooperation, proposals for norms, and next steps.[30] These are important institutional developments, but Assembly resolutions and exchanges do not themselves constitute a new treaty.

14.2 LAWS: CCW Negotiations Continue

The first 2026 session of the Group of Governmental Experts on Lethal Autonomous Weapons Systems under the CCW took place on March 2–6. The second is scheduled for August 31–September 4 and had not begun at this article's evidence cutoff.[33] The article therefore cannot present a “rolling text” still under negotiation as final legal norms already agreed. On August 25, 2026, the UN Secretary-General and ICRC President again publicly called for negotiations on a legally binding international instrument to begin as soon as possible. That appeal expresses a strong normative and policy position, but does not establish that a treaty already exists.[35]

14.3 REAIM and National / Alliance Policies: From Principles to Practice

REAIM 2026's Pathways to Action explicitly recognizes both military AI's potential benefits for personnel protection, civilian protection, and better decision-making, and its risks of miscalculation, bias, loss of control, and escalation. It identifies practical directions including legal reviews, risk assessments, TEVV, human involvement, and confidence-building measures.[31] UNIDIR's synthesis of five regional consultations in 2025 also shows that national policies, regional priorities, and stakeholder concerns are not identical. This further reminds us that REAIM is better understood as an open political process creating a shared space for action, rather than globally converged rules.[32] NATO, the United States, and other states and alliances have also developed their own responsible-use principles and review and testing requirements.[36–39] Whether these practices can become stabilizing mechanisms that are shared across opposing camps, verifiable, and do not create unilateral competitive disadvantages remains a question for international relations.

Governance layer	What has been established	What it must not be presented as
UNGA 79/239 and 80/58	A global political / institutional agenda; affirmation of international law's applicability and promotion of reports and exchanges	General Assembly resolutions, not a dedicated military AI treaty
UNGA 80/23 AI and nuclear C3	Dedicated treatment of possible risks of AI entering nuclear weapons command, control, and communications, with a call to maintain human control and oversight	Adopted 118–9–44; important political and normative pressure, not treaty obligations or unanimous practice
A/80/78 and the 2026 informal exchanges	Collection of national and other perspectives, forming a shared problem space	Diverse views, not a single consensus conclusion
CCW GGE on LAWS	Continuing negotiations on norms and operational frameworks for LAWS	Still under negotiation at the cutoff, with no comprehensive final instrument
REAIM Pathways to Action	State-led, multistakeholder political and practical recommendations	A political document open to endorsement, not universal legal obligations

Governance layer	What has been established	What it must not be presented as
NATO / national policies	Rules on responsible AI, testing, authorization, and use within alliances or states	Limited applicable actors; not globally unified norms
ICRC positions / joint appeals	Interpretations of humanitarian law, proposals for prohibitions and restrictions, and advocacy for legal development	Institutional positions and policy initiatives, not existing universal treaty provisions

Table 8. Layers of military AI governance as of August 26, 2026

15. The Boundary of Shared Survival | Where the Civilizational Perspective Enters and Where It Stops

Strategic studies examines how mutual competition produces security or insecurity; it does not independently decide “what kind of civilization deserves to exist.” Only here does the article introduce a category F civilizational floor: in scenarios that may produce highly irreversible consequences, such as nuclear escalation, uncontrolled autonomous weapons, and large-scale attacks on critical infrastructure, shared survival must be made visible as an independent consequence beyond localized national gains.

15.1 Shared Survival Does Not Cancel National Security

A state facing real threats has ample reason to defend itself, honor commitments to allies, and protect its people. Requiring it to accept unilateral, irreversible vulnerability for an abstract “humanity as a whole” does not automatically produce stable order. The more precise strategic meaning of shared survival is that parties' security arrangements must not treat “everyone moving closer to destruction” as a negligible externality. Mechanisms should be sought that make actors' long-term interests more compatible with the conditions for shared survival.

15.2 Shared Survival Does Not Generate World Sovereignty

“Some risks cross borders and may be irreversible” does not directly imply that “a single world government must be established.” Sovereignty, authorization, and legitimacy still require separate demonstration by political science, political philosophy, and law. International relations can establish at most that some externalities cannot be adequately internalized by unilateral security logic and therefore require coordination, confidence-building measures, arms control, or shared rules. Whether to use treaties, multilateral mechanisms, alliance arrangements, polycentric governance, or other structures must be decided by comparing their practical effects.

15.3 Minimal Category F Directions for Highly Irreversible Decisions

This article publicly assumes four commitments: responsibility must be traceable; authorization must be explicit; critical assumptions should be reviewable wherever possible; and pathways for correction, termination, and de-escalation must be preserved. These are not empirical facts of international relations, nor do they automatically have legal force. They follow from the possibility that irreversible consequences make “knowing afterward that we were wrong” meaningless. Any specific institution must still undergo joint validation for strategic feasibility, international law, domestic politics, and technical engineering.

16. What Evidence Could Change This Article | The Strongest Objections, Defeat Conditions, and Research Agenda

16.1 Strongest Objection One: Technology Is Not an Independent Cause of Escalation

This objection is substantially valid. Talmadge's historical research explicitly questions new technology as an independent, primary driver of escalation, emphasizing that technology is often a necessary but rarely sufficient intervening variable.[14] Accordingly, this article does not claim that AI technology determines war. If future research further demonstrates that political intentions, existing strategies, and organizational processes explain

almost all purported AI effects, the AI variable should be downgraded, and a separate professional paper may no longer be necessary.

16.2 Strongest Objection Two: AI May Improve Crisis Stability Overall

This is entirely possible. If better sensing, prediction, cyber defense, simulation, and communications significantly reduce false alarms over time and expand the options visible to leaders without creating new preemptive pressure or diluting responsibility, this article's claim that "speed competition compresses correction windows" must be narrowed. Shandler, Gross, and Shereshevsky conducted two experiments with 2,015 current or former Israeli military personnel using a high-fidelity military decision-support interface. Labeling the source as AI did not significantly shorten decision time in either study. After explanatory information was added in the second study, decision times still did not differ significantly between the AI and human-analyst groups.[45] This important task-level negative result is sufficient to reject the unconditional inference that "AI involvement necessarily makes people decide faster," but cannot alone adjudicate multiround interactions in interstate crises. A strategic security paper must not reject positive evidence, negative findings, or null results simply because risk narratives attract more attention.

16.3 Strongest Objection Three: Human-Machine Collaboration Can Reliably Solve the Control Problem

If long-term extrapolation from operational experience, exercises, organizational experiments, and accident data shows that highly automated systems consistently retain effective takeover, clear responsibility, and high-quality political judgment, without relegating humans to final-stage confirmation, then the risk that "a human in the loop does not mean human control" requires a narrower scope. The article is concerned with whether control is real, not with preserving a principle for its own sake.

16.4 Strongest Objection Four: International Governance Will Mature Faster Than Risks Diffuse

Current multilateral processes are indeed accelerating. If rules emerge across major military powers that are enforceable, verifiable, and capable of handling technical updates and crisis information, and consistently reduce miscalculation and competitive acceleration, this article's judgment that governance lags behind must be revised. Conversely, if rules work only in low-competition settings, their boundaries of applicability must also be precisely specified.

16.5 Defeat Conditions

Testable proposition	Facts that could overturn / change the article	Required revision
P3a Changed perceptions	Experiments and cross-national studies show that recognizing AI involvement or process changes does not systematically alter adversaries' judgments of threats, credibility, accidents, or responsibility	Remove or narrow the "changed perceptions" mechanism
P3b Direction of response	Adversary responses have no stable direction, and acceleration, preauthorization, or catching up are no more common than backing down, verification, oversight, refusal to use AI, or shared guardrails	Abandon the tendency toward symmetric acceleration and retain an open set of responses
P3c Joint time compression	Bilateral / multiround data reveal no reciprocal process feedback, or AI involvement does not jointly compress verification, communication, and political re-entry time	Remove or substantially narrow the "joint time compression" proposition

Testable proposition	Facts that could overturn / change the article	Required revision
P3d Transmission to crisis risk	Even when shared time budgets decline, risks of misperception, accidents, or unintended escalation do not increase, or are consistently offset by independent review, hotlines, reversible authorization, and similar measures	Narrow the risk-transmission chain and increase the weight of stabilizing mechanisms
Automation and responsibility	Long-term data show that greater automation does not dilute responsibility and that takeover quality consistently improves	Narrow organizational-control risks
Perception and misperception	Better situational awareness significantly reduces misperception, preemptive pressure, and mistaken attribution across different crises	Increase the weight of stabilizing effects
Strategic capability gaps	Capability differences after AI diffusion are much smaller than expected, without generating new alliance / national competition	Narrow judgments concerning relative advantage and competitive acceleration
Institutional guardrails	Transnational guardrails consistently offset new risks without creating unacceptable unilateral vulnerability	Weaken the judgment of inadequate governance
Independent factual domain	Political science, AI science, systems science, and international law can address all key propositions without exceeding their authority	Reassess or even withdraw the justification for a standalone article

Table 9. Defeat conditions and mandatory revision rules

16.6 Priority Research Agenda

- P3a | Adversary perception: compare identical information attributed to humans, AI, and different degrees of automation; test changes in judgments of threat, credibility, accidents, responsibility, and willingness to act. Preregister directional hypotheses and retain findings of no significant difference.
- P3b | Direction of response: treat retaliation, backing down, independent verification, catching up, countermeasures, oversight, refusal to use AI, and shared guardrails as parallel outcomes. Test which task, organizational, alliance, and crisis conditions promote each response.
- P3c | Shared time budget: measure T_V, T_C, and T_P separately in bilateral / multilateral, multiround wargames, observing whether one party's process changes trigger an adversary response that in turn changes the former. Report the three-dimensional vector, do not construct an overall score, and do not substitute unilateral or single-round acceleration for interaction evidence.
- P3d | Crisis consequences: compare whether joint time compression increases misperception, accidents, or unintended escalation across conditions of second-strike confidence, doctrinal transparency, independent verification, hotlines, and reversible authorization. Require reporting of opposite-direction, negative, and null results.
- Signaling and attribution: test how synthetic content, data contamination, and model-assisted attribution change leaders' confidence and escalation judgments at different levels of intelligence quality.
- Alliance correlation: study whether failures in shared models, cloud services, and data interfaces produce synchronized errors across allies, and whether diversification and redundancy mitigate them.
- Nuclear domain: develop research on transparency and confidence-building measures for AI entering early warning, NC3 assistance, and related conventional systems, without exposing sensitive operational details.

- Positive capabilities: systematically identify AI applications that consistently improve defense, accident detection, crisis communication, and de-escalation without significantly increasing adversaries' fears of preemption.
- Institutional experiments: compare the effects of legal reviews, TEVV, independent red teams, decision pauses, dual confirmation, model diversification, and transnational accident notification within real organizations.

Conclusion | When Greater Speed Becomes Possible, Slower Verification Becomes More Valuable

International relations and strategic security in the AGI era should not be reduced to “will AI start wars?” The real questions arise earlier: how states perceive threats anew through AI, change organizational processes for speed advantages, interpret each other's new capabilities, preauthorize more steps, share models and responsibility within alliances, and preserve time for reconsideration during crises.

Traditional strategic studies has not thereby become obsolete. Research on the security dilemma, deterrence, commitments, signaling, misperception, alliances, organizational accidents, and escalation provides precisely the professional legacy needed to avoid technological determinism. The new task in the AI era is to retest these established questions under new capability conditions: does improved performance remain improved security once it enters interaction? Does a state's locally rational choice still increase common security when aggregated with others' choices?

This article does not ask states to abandon competition or turn shared survival into a license for world sovereignty. It insists only on one floor: when errors may be irreversible, speed, efficiency, and capability cannot spare humans the need for verification, responsibility, communication, and de-escalation. The strategic value of stronger intelligence truly worth pursuing is not making war happen faster, but making more misperceptions visible before they become war.

When every party uses AI to make itself “more secure,” the shared world may become less secure. Strategic studies is responsible for identifying the conditions under which this statement holds, and those under which it does not. The latter are equally part of the answer.

Further Reading and Joint Research Entry Points

At the international relations / strategic security interface, this article answers only how AI entering interactions among states and alliances changes the security dilemma, deterrence, crisis stability, signaling, escalation, and risks to shared survival. It does not settle judgments of model safety, legal validity, political legitimacy, ecological carrying capacity, or civilization's ultimate direction. Further reading is therefore not a stack of documents, but an invitation for the next discipline with genuine factual responsibility to re-enter.

Upward | Why Does Strategic Security Ultimately Become a Civilizational Question?

Proceed to [Reckoning with the Disciplines in the Age of AGI: Disciplinary Boundaries, Interactions, and a Shared Research Agenda V2.4](#), and [Opportunities, Directions, Methods, and Goal Realization for Humanity's Civilization Leap in the Age of AGI](#), the convergent research paper on civilizational direction V1.1. The former explains why disciplines must remain independent and why real-world causality cannot stop at their boundaries. The latter publicly assumes responsibility for judging where these shared changes ought to take humanity. Strategic stability is one condition for survival, not civilization's entire purpose.

Downward | To Continue Within International Relations and Strategic Security

Proceed into deterrence theory, crisis history, arms control, the security dilemma, alliances, intelligence and misperception, nuclear strategy, cyber conflict, command and control, wargaming, and cross-national crisis databases. The priority is not another sweeping judgment, but an actual research design comparing purely human processes, different forms of AI assistance, and different automation and preauthorization structures.

Leftward | Which Disciplines Must Re-Enter?

Discipline re-entering	Why this article cannot substitute for it
AI / computer science	What systems can do, how they fail, how to test them, and whether permissions and stopping mechanisms are reliable
Political science	The legitimacy of state capacity, public power, sovereignty, and authorization
Ethics / law / international law	Use of force, international humanitarian law, responsibility, rights, and boundaries of harm
Systems science	General mechanisms of cascading effects, tight coupling, correlated failures, and resilience
Epistemology / cognitive science	Intelligence uncertainty, model inference, automation dependence, time pressure, and misperception
Economics / political economy	Arms investment, industrial structure, supply chains, compute, and strategic resources
Sociology / anthropology	Friend–enemy narratives, trust, identity, mobilization, and the social consequences of war

Table 10. Principal disciplinary handoffs of the strategic studies paper

Rightward | If You Disagree, Begin with the Strongest Opposition

Prioritize four lines of inquiry capable of changing this article: whether AI strengthens strategic stability overall; whether new technology is only an intervening variable in existing political conflicts rather than an independent driver; whether stronger human-machine collaboration can resolve control and responsibility problems; and whether difficulties of verification, relative advantage, and dual use impede military AI arms control. Contrary evidence is not an appendix, but a condition of the article's continued validity.

Forward | What Professional Work Could Change This Article?

New crisis data, multinational exercises, organizational experiments, wargaming, negative findings, accident materials, formal models, signaling experiments, institutional comparisons, arms-control verification tools, and red-team results can all directly change this article. If [Human-COS](#), a candidate system for multi-actor, multiperspective consequence simulation and judgment support before high-consequence choices become irreversible, is used in strategic research, it can serve only as a candidate research and validation interface. It cannot in reverse become evidence for strategic facts.

The Civilizational Future Chain | After Avoiding Destruction, Where Else Does Humanity Want to Go?

To explore how human work, education, families, relationships, public life, and technological collaboration might change when strategic security no longer consumes all of civilization's capacities over the long term and technology does more to protect than compress a shared future, proceed to the [Living Civilization trilogy](#) and the [five-volume series](#). One boundary must remain: vision is not evidence. Living Civilization enables readers to see, compare, reject, or reconstruct possible new worlds. It does not prove any strategic causality in this article.

Civilization Leap five-volume series: <https://www.civitas.top/en/documents/p02-five-001/> | Public entry point: <https://www.civitas.top/en/documents/p01-entry-006/> | Complete library: <https://www.civitas.top/en/library/>

The most valuable participation is not agreement, but professional work capable of changing the theory.

The purpose is not to get more people to finish reading Civilization Leap, but to bring more people together to complete Civilization Leap.

Appendix A | Register of Source–Claim Boundaries

Source type	How this article uses it	What it explicitly cannot establish
Classical research on the security dilemma and deterrence	Explains why interaction, commitment, offense and defense, and signaling matter	Cannot directly establish that the AI era will replicate past outcomes
Academic research on emerging technologies / AI strategy	Proposes competing mechanisms concerning speed, autonomy, organization, offense and defense, and stability	Mostly theory, historical analogy, or limited cases; cannot be treated as future facts
SIPRI / UNIDIR research	Integrates public sources and provides regional perspectives and baselines for military tasks and nuclear risk	Institutional reports do not reveal states' actual secret procedures or constitute international legal consensus
UNGA / UNODA	Establishes the status of the global agenda, formal resolutions, reports, and meetings	Assembly resolutions / meetings do not automatically constitute a new treaty
UNGA 80/23 (AI and nuclear C3)	Establishes that the General Assembly has made possible risks of integrating AI into nuclear weapons command, control, and communications a distinct formal agenda item, and records its calls for human control / oversight and voting status	Cannot establish uniform practice among nuclear-weapon states, universal treaty obligations, or that AI has entered final authorization of nuclear use
REAIM / NATO / US policies	Establishes responsible AI principles and governance practices of states or alliances	Political commitments and national policies do not represent globally unified norms
ICRC	Establishes its interpretations of humanitarian law, policy positions, and proposals for new rules	ICRC positions cannot be presented as existing treaty rules accepted by all states

Table A1. Boundaries of sources and claim strength

Appendix B | Strategic Stress-Test Matrix

Stress event	First factual layer engaged	Strategic question	Priority checks
An early-warning model makes a high-confidence error	Data / model / human calibration issues	Crisis stability: does it trigger preemptive / retaliatory pressure?	Possibility of independent review, delay, communication, and de-escalation
One side deploys faster automated defense	Engineering improves response	Security dilemma: how does the other side interpret its offensive / defensive character?	Posture distinguishability and crisis communication
Allies share the same foundation model	Interoperability improves efficiency	Correlated failures and shared responsibility	Model diversity, backups, version visibility
A deepfake simulates a national leader's orders	Information authenticity is compromised	Signaling and attribution: is political intent misperceived?	Source authentication, human callbacks, multichannel verification

Stress event	First factual layer engaged	Strategic question	Priority checks
AI-assisted conventional counterforce strikes	Improved target identification	Perceptions of nuclear vulnerability and cross-domain escalation	Strategic dialogue, limiting misinterpretation, command-and-control resilience
A supplier's service is interrupted	Commercial / technical failure	State or alliance strategic dependence is exposed	Alternative supply, low-tech backups, sovereignty boundaries
A nonstate actor uses a general-purpose model	Task barriers are lowered	Uncertainty over attribution and the target of retaliation	Task-level capability assessment, evidentiary chains, and proportionate responses
AI advice substantially improves crisis decisions	Positive performance emerges	Do this article's risk claims need narrowing?	Replication across cases, capabilities after removing AI, and adversary responses

Table B1. Eight types of interdisciplinary strategic stress tests

Declaration of Generative AI Use

Generative artificial intelligence was used in researching and writing this manuscript to assist with organizing materials and bibliographic leads, structural design, drafting and polishing some text, cross-checking, verifying citation leads, and checking layout. AI outputs are not empirical data, authoritative sources, or independent evidence. All key facts, literature, formal documents, and time-sensitive information must be traced back to verifiable sources. The author reviews and approves the research questions, theoretical propositions, A/B/C/F status judgments, selection of materials, normative positions, adoption of citations, and final manuscript, and assumes final academic responsibility. Generative AI is not listed as an author. At formal submission, the tools, model versions, dates of use, and specific scope will be disclosed further in accordance with the target journal's requirements.

References

- [1] Baldwin, David A. "The Concept of Security." *Review of International Studies* 23, no. 1 (1997): 5–26.
<https://doi.org/10.1017/S0260210597000053>.
- [2] Herz, John H. "Idealist Internationalism and the Security Dilemma." *World Politics* 2, no. 2 (1950): 157–180.
<https://doi.org/10.2307/2009187>.
- [3] Jervis, Robert. "Cooperation under the Security Dilemma." *World Politics* 30, no. 2 (1978): 167–214.
<https://doi.org/10.2307/2009958>.
- [4] Snyder, Glenn H. "The Security Dilemma in Alliance Politics." *World Politics* 36, no. 4 (1984): 461–495.
<https://doi.org/10.2307/2010183>.
- [5] Fearon, James D. "Rationalist Explanations for War." *International Organization* 49, no. 3 (1995): 379–414.
<https://doi.org/10.1017/S0020818300033324>.
- [6] Huth, Paul, and Bruce Russett. "What Makes Deterrence Work? Cases from 1900 to 1980." *World Politics* 36, no. 4 (1984): 496–526.
<https://doi.org/10.2307/2010184>.
- [7] Schelling, Thomas C. *Arms and Influence*. New Haven: Yale University Press, 1966.
- [8] Jervis, Robert. *Perception and Misperception in International Politics*. Princeton: Princeton University Press, 1976.
- [9] Allison, Graham T., and Philip Zelikow. *Essence of Decision: Explaining the Cuban Missile Crisis*. 2nd ed. New York: Longman, 1999.
- [10] Posen, Barry R. *Inadvertent Escalation: Conventional War and Nuclear Risks*. Ithaca: Cornell University Press, 1991.
- [11] Sagan, Scott D. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton: Princeton University Press, 1993.

- [12] Perrow, Charles. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books, 1984.
- [13] Sechser, Todd S., Neil Narang, and Caitlin Talmadge. "Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War." *Journal of Strategic Studies* 42, no. 6 (2019): 727–735. <https://doi.org/10.1080/01402390.2019.1626725>.
- [14] Talmadge, Caitlin. "Emerging Technology and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today." *Journal of Strategic Studies* 42, no. 6 (2019): 864–887. <https://doi.org/10.1080/01402390.2019.1631811>.
- [15] Garfinkel, Ben, and Allan Dafoe. "How Does the Offense-Defense Balance Scale?" *Journal of Strategic Studies* 42, no. 6 (2019): 736–763. <https://doi.org/10.1080/01402390.2019.1631810>.
- [16] Horowitz, Michael C. "Artificial Intelligence, International Competition, and the Balance of Power." *Texas National Security Review* 1, no. 3 (2018): 36–57. <https://doi.org/10.15781/T2639KP49>.
- [17] Horowitz, Michael C. "When Speed Kills: Lethal Autonomous Weapon Systems, Deterrence and Stability." *Journal of Strategic Studies* 42, no. 6 (2019): 764–788. <https://doi.org/10.1080/01402390.2019.1621174>.
- [18] Altmann, Jürgen, and Frank Sauer. "Autonomous Weapon Systems and Strategic Stability." *Survival* 59, no. 5 (2017): 117–142. <https://doi.org/10.1080/00396338.2017.1375263>.
- [19] Goldfarb, Avi, and Jon R. Lindsay. "Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War." *International Security* 46, no. 3 (Winter 2021/22): 7–50. https://doi.org/10.1162/isec_a_00425.
- [20] Nadibaidze, Anna, and Nicolò Miotto. "The Impact of AI on Strategic Stability is What States Make of It: Comparing US and Russian Discourses." *Journal for Peace and Nuclear Disarmament* 6, no. 1 (2023): 47–67. <https://doi.org/10.1080/25751654.2023.2205552>.
- [21] Erskine, Toni, and Steven E. Miller. "AI and the Decision to Go to War: Future Risks and Opportunities." *Australian Journal of International Affairs* 78, no. 2 (2024): 135–147. <https://doi.org/10.1080/10357718.2024.2349598>.
- [22] Boulanin, Vincent, Lora Saalman, Petr Topychkanov, Fei Su, and Moa Peldán Carlsson. *Artificial Intelligence, Strategic Stability and Nuclear Risk*. Stockholm: SIPRI, June 2020. <https://www.sipri.org/publications/2020/policy-reports/artificial-intelligence-strategic-stability-and-nuclear-risk>.
- [23] Chernavskikh, Vladislav. *Nuclear Weapons and Artificial Intelligence: Technological Promises and Practical Realities*. Stockholm: SIPRI, September 2024. <https://doi.org/10.55163/VBQX6088>.
- [24] Saalman, Lora. *AI in Chinese, Indian and US Nuclear Postures, Norms and Systems*. Stockholm: SIPRI, February 2026. <https://doi.org/10.55163/QZBL7397>.
- [25] Grand-Clément, Sarah. *Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain*. Geneva: UNIDIR, 2023. <https://unidir.org/publication/artificial-intelligence-beyond-weapons-application-and-impact-of-ai-in-the-military-domain/>.
- [26] Unal, Beyza, and Ulysse Richard. *Governance of Artificial Intelligence in the Military Domain*. UNODA Occasional Papers No. 42. New York: United Nations, June 2024. ISBN 978-92-1-003174-5.
- [27] United Nations General Assembly. Resolution 79/239, "Artificial intelligence in the military domain and its implications for international peace and security." Adopted 24 December 2024. A/RES/79/239. <https://docs.un.org/A/RES/79/239>.
- [28] United Nations Secretary-General. *Artificial intelligence in the military domain and its implications for international peace and security: Report of the Secretary-General*. A/80/78, 5 June 2025. <https://digitallibrary.un.org/record/4086346>.
- [29] United Nations General Assembly. Resolution 80/58, "Artificial intelligence in the military domain and its implications for international peace and security." Adopted 1 December 2025. A/RES/80/58. <https://docs.un.org/A/RES/80/58>.
- [30] United Nations Office for Disarmament Affairs. "Informal exchanges on artificial intelligence in the military domain." Geneva, 15–17 June 2026, convened pursuant to A/RES/80/58. <https://meetings.unoda.org/unoda-stu-meeting/unoda-science-technology-and-international-security-unit-meeting-2026>.
- [31] REAIM 2026. *A Coruña Declaration: "REAIM 2026 Pathways to Action."* 5 February 2026. <https://www.reaim2026.es/en/REAIM2026/Documents/REAIM%202026%20Pathways%20to%20Action.pdf>.
- [32] Afina, Yasmin. *The Global Prism of Military AI Governance: Reflections from the 2025 Regional Consultations on Responsible AI in the Military Domain*. Geneva: UNIDIR, 2 February 2026. <https://unidir.org/publication/the-global-prism-of-military-ai-governance-reflections-from-the-2025-regional-consultations-on-responsible-ai-in-the-military-domain/>.
- [33] United Nations Office for Disarmament Affairs. *Convention on Certain Conventional Weapons—Group of Governmental Experts on Lethal Autonomous Weapons Systems (2026)*, first session 2–6 March 2026 and second session scheduled 31 August–4 September 2026. <https://meetings.unoda.org/ccw-/convention-on-certain-conventional-weapons-group-of-governmental-experts-on-lethal-autonomous-weapons-systems-2026>.

- [34] International Committee of the Red Cross. Autonomous Weapon Systems and International Humanitarian Law: Selected Issues. Geneva: ICRC, 5 January 2026. <https://www.icrc.org/en/publication/autonomous-weapon-systems-and-international-humanitarian-law-selected-issues>.
- [35] United Nations Secretary-General and International Committee of the Red Cross President. "Renewed call from UN Secretary-General and ICRC President to adopt rules on autonomous weapon systems." Joint appeal, 25 August 2026. <https://www.icrc.org/en/statement/renewed-call-un-secretary-general-and-icrc-president-adopt-rules-autonomous-weapons>.
- [36] NATO. "Summary of NATO's revised Artificial Intelligence (AI) strategy." 10 July 2024. <https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy>.
- [37] U.S. Department of Defense. DoD Directive 3000.09, Autonomy in Weapon Systems. 25 January 2023. <https://media.defense.gov/2023/Jan/25/2003149928/-1/-1/0/DODDIRECTIVE-3000.09-AUTONOMY-IN-WEAPON-SYSTEMS.PDF>.
- [38] U.S. Department of State. Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy. Latest version, 1 November 2023. Archived official text: <https://2021-2025.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy>.
- [39] NATO. The Secretary General's Annual Report 2025. Brussels: NATO, 2026. <https://www.nato.int/content/dam/nato/webready/documents/publications-and-reports/annual-reports/sgar25-en.pdf>.
- [40] Gartzke, Erik. "The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth." *International Security* 38, no. 2 (2013): 41–73. https://doi.org/10.1162/ISEC_a_00136.
- [41] Lindsay, Jon R. "Stuxnet and the Limits of Cyber Warfare." *Security Studies* 22, no. 3 (2013): 365–404. <https://doi.org/10.1080/09636412.2013.816122>.
- [42] United Nations General Assembly. Resolution 80/23, "Possible risks of the integration of artificial intelligence into command, control and communications systems of nuclear weapons." Adopted 1 December 2025. A/RES/80/23. <https://docs.un.org/A/RES/80/23>.
- [43] Horowitz, Michael C., and Erik Lin-Greenberg. "Algorithms and Influence: Artificial Intelligence and Crisis Decision-Making." *International Studies Quarterly* 66, no. 4 (2022): sqac069. <https://doi.org/10.1093/isq/sqac069>.
- [44] Zeng, Jinghan. "The US Factor in Chinese Perceptions of Militarized Artificial Intelligence." *International Affairs* 101, no. 2 (2025): 677–689. <https://doi.org/10.1093/ia/iaae323>.
- [45] Shandler, Ryan, Michael L. Gross, and Yahli Shereshevsky. "Black Box Warfare: Human Judgment and Military Decision-Making in the Age of AI." *Journal of Conflict Resolution*, OnlineFirst (2026). <https://doi.org/10.1177/00220027261463443>.
- [46] Lushenko, Paul. "AI, Trust, and the War Room: Evidence from a Conjoint Experiment in the US Military." *Cambridge Forum on AI: Law and Governance* 1 (2025): e52. <https://doi.org/10.1017/cfl.2025.10019>.
- [47] Schwartz, Joshua A., and Michael C. Horowitz. "Delegating Destruction: Coercive Threats and Automated Nuclear Systems." *International Organization* 80, no. 1 (2026): 179–204. <https://doi.org/10.1017/S0020818325101215>.
- [48] Johnson, James. "Revisiting the 'Stability–Instability Paradox' in AI-Enabled Warfare: A Modern-Day Promethean Tragedy under the Nuclear Shadow?" *Review of International Studies*, First View (2024): 1–19. <https://doi.org/10.1017/S0260210524000767>.
- [49] Zatssepina, Luba. "Waltzing into Uncertainty: AI in Nuclear Decision Making and the Challenge of Divergent Deterrence Logics." *Cambridge Forum on AI: Law and Governance* 1 (2025): e56. <https://doi.org/10.1017/cfl.2025.10021>.
- [50] Johnson, James. "GenAI and Synthetic Foresight at the Brink: The Future of Nuclear Crisis Decision-Making." *European Journal of International Security*, First View (2026): 1–22. <https://doi.org/10.1017/eis.2026.10063>.
- [51] Yuan, Shaoyu. "World and War in the Age of AI: Cyber Risk, Deterrence, and Decision-Making in U.S.-China Competition." *Defence Studies*. Published online 15 August 2026. <https://doi.org/10.1080/14702436.2026.2718345>.

Version Note

Version note: V1.2.1 is a minimal citation patch. Building on V1.2, it adds Yuan (2026) to the references and a paragraph specifying the boundaries of originality. It clarifies that this article does not claim interactive time compression has been empirically observed, nor that it first raised the broad questions of action–reaction pressures, adoption incentives, deterrence, and crisis stability. Its specific contribution remains limited to the separation of P3a–P3d into independently defeasible propositions, the shared time budget (T_V , T_C , T_P), and the corresponding failure conditions. V1.2's principal argument, disciplinary object, A/B/C/F boundaries, category F positioning of shared survival, and interdisciplinary handoffs remain unchanged. This version update does not change the theoretical core, elevate evidentiary status, or constitute an English-language external-circulation or submission manuscript. V1.2.1 is the current Canonical professional research baseline.